

Experimente mit stochastischen Automaten

Der Technischen Fakultät
der Universität Erlangen-Nürnberg
zur Erlangung des Grades
DOKTOR - INGENIEUR

vorgelegt

von

Johann Friedrich Böhme

Erlangen 1970

1970 # 392

Experimente mit stochastischen Automaten

Der Technischen Fakultät
der Universität Erlangen-Nürnberg
zur Erlangung des Grades
DOKTOR - INGENIEUR



vorgelegt

von

Johann Friedrich Böhme

Erlangen 1970



Als Dissertation genehmigt von der Technischen
Fakultät der Universität Erlangen-Nürnberg

Tag der Einreichung: 17. 5.1970

Tag der Promotion: 27. 7.1970

Dekan: Prof. Dr. W. Schüßler

Berichterstatter: Prof. Dr. W. Händler

Prof. Dr. B. Dejon

Zusammenfassung

1.	Einleitung	1
1.1	Formen von Experimenten mit stochastischen Automaten	1
1.2	Erläuterung einiger behandelte Probleme	4
2.	n-fach Experimente mit diskreten Kanälen ohne Vorgriff	8
2.1	Allgemeine Theorie	8
2.1.1	Definitionen	8
2.1.2	Voraussetzungen	15
2.1.3	Aussagen	21
2.2	Spezialisierungen durch Wahl der Kostenfunktion	33
2.2.1	Buchstabenweise adaptive Experimente	34
2.2.2	Nichtadaptive Experimente	36
2.3	Vorgabeexperimente	41
2.4	Vorgabeexperimente bei unvollständiger Kenntnis des Experimentators über das Experiment	46
2.5	Bemerkungen	48
3.	Einige Aussagen über zustandsbestimmende Vorgabeexperimente mit stochastischen Automaten	54
3.1	Schranken für den Entscheidungsfehler bei n-fach diagnostischen Vorgabeexperimenten	55
3.1.1	Obere Schranken	56
3.1.2	Untere Schranken für große n	61
3.1.3	Untere Schranken für kleine n	72
3.2	Schranken für den Entscheidungsfehler bei einfachen heimführenden Vorgabeexperimenten	79
3.3	Bemerkungen	83
	Verzeichnis der wichtigsten Definitionen	88
	Literaturverzeichnis	90

Zusammenfassung

Nach einem einleitenden ersten Abschnitt wird im zweiten Teil dieser Arbeit eine Theorie der Experimente mit einem gestörten Kanal entwickelt. Die Kanäle sollen sequentiell arbeiten, endliche Ein- und Ausgabealphabete, keinen Vorgriff und im allgemeinen ein unendlich langes Gedächtnis besitzen. Diese Theorie stellt auch eine allgemeine Theorie für Experimente mit einem stochastischen Automaten dar. Dabei werden die Experimentiervorschriften, nach denen die Experimente auszuführen sind, als statistische Entscheidungsfunktionen betrachtet und damit eine statistische Entscheidungstheorie aufgebaut. Der dritte Teil befaßt sich mit diagnostischen und heimführenden Vorgabeexperimenten mit einem stochastischen Automaten, für die im wesentlichen Abschätzungen für die Entscheidungsfehler angegeben werden.

Abstract

This paper presents a theory of experiments with a noisy channel. The channels are sequential and non anticipating, and they have finite alphabets for input and output and an infinite memory. With this we have a general theory of experiments with a stochastic automaton. Therein instructions for experimentations are considered as statistical decision functions. In the last part we are concerned with diagnosing and homing preset experiments with a stochastic automaton. Essentially it contains estimations of decision errors.

1. Einleitung

Der erste Abschnitt dieser Arbeit ist eine breite Einführung in die Probleme, die in den Kapiteln 2 und 3 behandelt werden. Die Ausführlichkeit ist notwendig, da an anderer Stelle noch nicht über Experimente mit stochastischen Automaten in allgemeiner Form publiziert worden ist. Im ersten Teil der Einleitung wird eine bekannte Definition der endlichen stochastischen Automaten angegeben und erläutert, was ein Experiment mit einem endlichen stochastischen Automaten sein soll. Anschließend werden denkbare Formen von solchen Experimenten aufgezählt. Im zweiten Teil wird verbal gezeigt, wie man alle diese Formen in einer Sicht betrachten kann. Dadurch werden Experimente mit Kanälen motiviert. Eine gründliche Behandlung der Theorie der Experimente mit diskreten Kanälen ohne Vorgriff findet der Leser in Kapitel 2.

1.1 Formen von Experimenten mit stochastischen Automaten

Ein endlicher stochastischer Automat (ab jetzt Automat genannt) wird nach Carlyle [8] durch ein Quadrupel (X, Y, Z, p) beschrieben, wobei X, Y und Z endliche Mengen sind, genannt Eingabealphabet, Ausgabealphabet und Menge der Zustände, sowie p ein Funktional ist mit

$$p: Z \times X \rightarrow \mathcal{M}^1(Y \times Z).$$

Hierbei bedeutet $\mathcal{M}^1(B)$ die Menge aller Wahrscheinlichkeitsmaße über der endlichen Menge B .

Wenn $x \in X, y \in Y, z \in Z$ und $z' \in Z$, so bezeichnet man mit $p(y, z' | z, x)$ die Wahrscheinlichkeit (Wk) des Paares (y, z') , wenn p an der Stelle (z, x) das vorliegende Wk-maß ist. In einem geeigneten Maßraum ist $p(y, z' | z, x)$ als bedingte Wk interpretierbar. Dann wird das Verhalten des

Automaten bei Vorlage einer diskreten Zeitskala wie folgt beschrieben: Befand sich der Automat im Zeitpunkt $t-1$ im Zustand z_{t-1} und wird zur Zeit t das Symbol x_t eingegeben, so ist die Wk dafür, daß der Automat im Zeitpunkt t gerade y_t ausgibt und sich im Zustand z_t befindet, $p(y_t, z_t | z_{t-1}, x_t)$.

Sei wie üblich B^* die Menge aller durch Konkatenation entstandener Wörter endlicher Länge über der endlichen Menge B und $l(u)$ die Länge eines Wortes u aus B^* . Λ sei das leere Wort und habe die Länge $l(\Lambda) = 0$.

Man erweitert die Definition von p bekanntermaßen:

$$1) \quad p(\Lambda, z | z, \Lambda) = 1,$$

$$2) \quad p(vy, z' | z, ux) = \sum_{z''} p(y, z' | z'', x) p(v, z'' | z, u)$$

für $z, z' \in Z$, $x \in X$, $y \in Y$, $u \in X^*$ und $v \in Y^*$ mit $l(v) = l(u)$.

Häufig liegt eine Anfangsverteilung q der Zustände des Automaten vor. D.h. der Automat arbeitet nur für positive Zeiten. Zur Zeit $t=0$ befindet sich der Automat mit der Wk $q(z)$ im Zustand z , wobei $\sum_{z \in Z} q(z) = 1$ gilt. Man bezeichnet gewöhnlich einen Automaten mit Anfangsverteilung durch ein Quintupel (X, Y, Z, p, q) .

Deterministische Automaten kann man als Spezialfälle von Automaten betrachten. Z.B. ist für deterministische Mealy-Automaten $(X, Y, Z, \delta, \lambda)$, wobei δ die Überföhrungs- und λ die Ausgabefunktion sind, der entsprechende Automat durch $p(\lambda(z, x), \delta(z, x) | z, x) = 1$ festgelegt.

Unter einem Experiment mit einem Automaten wird folgender Sachverhalt verstanden: Dem Experimentator (ER) liegen endlich viele voneinander unabhängig arbeitende Kopien eines Automaten in derselben Anfangssituation vor.

Der ER besitzt nicht alle Informationen über den Automaten (andernfalls wäre ein Experiment überflüssig) und möchte sich die fehlende Information verschaffen. Dazu hat der ER nur die Möglichkeit, einzelne Wörter in die Kopien des Automaten zu geben⁺) und die Ausgaben der Kopien zu lesen. Erschwerend kann noch hinzukommen, daß der ER nicht genau weiß, welche Buchstaben er eingegeben hat oder daß er gewisse Ausgabesymbole nicht lesen kann. Allein aus diesen Angaben und dem, was er schon über den Automaten weiß, kann er Schlußfolgerungen ziehen, z.B. weitere Wörter eingeben und die entsprechenden Reaktionen beobachten oder eine endgültige Aussage treffen. Im Gegensatz zum Experimentieren mit deterministischen Automaten sind die Reaktionen von zwei Kopien auf dieselbe Eingabe nicht notwendig gleich, obwohl dieselbe Anfangssituation vorgelegen hat. Die gewonnenen Aussagen des ER sind daher i.a. nicht fehlerfrei.

Man kann sich nun viele Formen von Experimenten mit Automaten denken und diese z.B. folgendermaßen sortieren: Zunächst wird man fragen, wieviele Kopien für ein Experiment benötigt werden. Ist es nur eine Kopie, so spricht man von einfachen, sonst von mehrfachen Experimenten. Weiterhin kann man Experimente nach der Vorgehensweise des ER einteilen: Wird nach höchstens einer Eingabe von Worten in die Kopien (nicht alle Wörter dürfen gleich sein) eine endgültige Aussage getroffen, so heißt das Experiment nicht adaptiv und sonst adaptiv. Kennt der ER a priori die Eingabewörter im nicht adaptiven Fall, so heißt das Experiment Vorgabeexperiment. Man unterscheidet auch, ob dem ER die volle Information über Eingaben und Ausgaben zur Verfügung steht oder nicht. Der zweite Fall wird in dieser Arbeit nur am Rande berührt werden.

⁺) d.h. nacheinander die Buchstaben der Wörter in die entsprechenden Kopien einzugeben

Schließlich ist es möglich, die Experimente nach ihren Zielsetzungen zu klassifizieren. U.a. findet man Aufgaben wie Zustandsidentifikation bei bekanntem bzw. nur teilweise bekanntem Funktional ρ . Hier sind zu nennen diagnostische Experimente, die also den Anfangszustand schätzen sollen, heimführende Experimente, welche den Endzustand bestimmen, oder auch synchronisierende Experimente, bei denen nach Eingabe von wohldefinierten Wörtern und ohne die Ausgabe zu berücksichtigen, klar ist, in welchem Endzustand sich der Automat befindet. Andere Experimente sollen bei bekanntem ρ die unbekannte Anfangsverteilung q schätzen (oder entsprechende Intervallabschätzungen). Schließlich sind noch die Automatenidentifikationsexperimente zu nennen. Darunter fällt z.B. das Problem zu entscheiden, ob ein vorgegebener Automat gewisse Normen erfüllt oder nicht. Anders ausgedrückt kann das auch heißen, ob ein Automat fehlerhaft arbeitet oder nicht. Man kann auch sagen, daß ein Automat aus einer gegebenen Klasse zu identifizieren ist oder zu entscheiden ist, ob der Automat zu einer gewissen Teilklasse gehört oder nicht. In Gill [11] findet man eine systematische Behandlung von Experimenten mit deterministischen Automaten. Dieses Buch kann als Anregung benutzt werden, weitere Experimente für Automaten zu definieren.

Nicht alle Experimente sind für stochastische Automaten sinnvoll, z.B. ein mehrfach heimführendes Vorgabeexperiment, da gleiche Eingaben in zwei Kopien nicht notwendig den gleichen Endzustand erzeugen.

1.2 Erläuterung einiger behandelte Probleme

Betrachtet man die Vielfalt von möglichen Experimenten mit Automaten, wie sie im letzten Abschnitt angedeutet worden ist, dann stellt sich zunächst das Problem, wie man alle diese Formen von Experimenten von einer Warte

aus betrachten könnte. Im folgenden wird eine Möglichkeit dafür angedeutet.

Zunächst sei bemerkt, daß man bei jedem Experiment den vorliegenden Automaten mit fester Anfangssituation als einen Kanal auffassen kann. Der Kanal hat dasselbe Ein- und dasselbe Ausgabealphabet wie der Automat. Er arbeitet sequentiell genau wie der Automat, d.h. auf jedes eingegebene Symbol des Eingabealphabets wird genau ein Symbol des Ausgabealphabets ausgegeben. Der Kanal hat keinen Vorgriff, d.h. später eingegebene Symbole haben keinen Einfluß auf früher ausgegebene Symbole. Er hat aber i.a. ein Gedächtnis, was bedeutet, daß die ausgegebenen Symbole von allen vorher eingegebenen Symbolen abhängig sind. Der Kanal hat schließlich dasselbe Eingabe-Ausgabe-Verhalten wie der Automat. Also ist stets die Wk dafür, daß das Wort v ausgegeben wird, wenn das Wort u mit $l(u) = l(v)$ eingegeben wird, für beide gleich.

Die Information, die der ER über den vorliegenden Automaten weiß und diejenige, die er nicht weiß, kann man so beschreiben: Der ER kennt nicht den vorliegenden Kanal, weiß aber, daß der Kanal aus einer dem ER bekannten Menge von Kanälen stammt. Die verschiedenen Aussagen, die der ER machen kann, werden jetzt zu einer abstrakten Aussagenmenge zusammengefaßt, die der ER für das durchzuführende Experiment kennt.

Ein Experiment mit einem Kanal wird nun genau analog zu einem Experiment mit einem Automaten durchgeführt. Ein Kanal mit Kopien, in die noch nichts eingegeben worden ist, aus einer bekannten Menge von Kanälen liegt dem ER vor. Der ER kann Eingaben in die Kanäle vornehmen, die Reaktionen beobachten, daraus Schlußfolgerungen ziehen und findet zuletzt eine Aussage aus der festen Aussagenmenge. Verallgemeinernd kann man annehmen, daß der ER

seine Entscheidungen zufällig trifft und daß die entsprechenden Wk-maße von den Ein- und Ausgaben der Kopien abhängen.

Man wird Verluste der Entscheidungen zu berücksichtigen haben. Das soll heißen, daß Fehlentscheidungen bei Vorlage eines bestimmten Kanales nicht notwendig gleich schwerwiegend sind. Andererseits müssen i.a. auch Kosten, z.B. die zur Ausführung des Experimentes benötigte Zeit, die wiederum von der Länge des Experimentes abhängt, beachtet werden. Als Maße für die Güte eines Experimentes könnte man dann erwartete Verluste und erwartete Kosten wählen. Ein wesentliches Problem besteht nun darin, zu definieren, was ein günstiges Experiment sein soll.

Das Experimentieren kann man auch als ein Zwei-Personen-Spiel auffassen (vergl. [15], [18]). Der erste Spieler ist die Natur oder irgendeine Person und der zweite der Experimentator. Eine Wahl des ersten Spielers für eine reine Strategie ist die Wahl des dem ER vorliegenden Kanales. Eine reine Strategie des ER wäre dann eine Vorschrift, nach der das Experiment auszuführen ist. Der Gewinn der Spieler müßte sich dann aus erwarteten Verlusten und den erwarteten Kosten zusammensetzen. Häufig kann man das Verhalten der Natur bei der Wahl des Kanales mit Hilfe einer gemischten Strategie beschreiben. Das bedeutet, daß es ein Wk-maß über der Menge der Kanäle gibt, nach dem die Natur zufällig ihre reine Strategie, nämlich den Kanal, wählt.

In Kapitel 2 wird eine den vorangegangenen Bemerkungen entsprechende Theorie von Experimenten mit Kanälen ausführlich behandelt. Dabei wird zunächst vorausgesetzt, daß dem ER alle Informationen über die Ein- und Ausgaben zur Verfügung stehen. Es werden Experimentiervorschriften (EV) definiert, nach denen ein Experiment durchgeführt

wird, und im wesentlichen eine Theorie der EV aufgebaut. Das Gütekriterium für eine EV wird die Summe aus erwarteten Verlusten und erwarteten Kosten sein. Als günstige EV werden Bayes-EV und Minimax-EV definiert. Nach der Festlegung von topologischen Strukturen in der Menge der Kanäle, der Menge der Aussagen und der Menge der EV werden dann Existenzsätze über günstige Lösungen bewiesen und die Beziehungen der EV untereinander beleuchtet. Schließlich werden noch einige allgemeine Aussagen über spezielle EV angefügt.

Das Kapitel 3 befaßt sich mit diagnostischen und heimföhrenden Vorgabeexperimenten mit einem stochastischen Automaten. Man erhält über diese Experimente mit Hilfe von Methoden aus der Informationstheorie sehr interessante spezielle Ergebnisse. Z.B. kann man Abschätzungen für den erwarteten Fehler beim Entscheiden (ein Spezialfall der erwarteten Verluste) angeben und das asymptotische Verhalten dieser Fehler in Abhängigkeit von der Zahl der Kopien des vorliegenden Automaten untersuchen.

Die Bemerkungen am Ende der Kapitel 2 und 3 sollen dazu dienen, die in den vorangehenden Abschnitten diskutierten Probleme noch etwas zu erläutern und einige Zusammenhänge mit Fragestellungen der Automatentheorie aufzuzeigen.

2 n-fach Experimente mit diskreten Kanälen ohne Vorgriff

Angeregt durch Arbeiten von A. Wald (siehe [18]) wird in diesem Kapitel eine Entscheidungstheorie für n-fach Experimente mit diskreten Kanälen ohne Vorgriff entwickelt. Die Theorie wird im wesentlichen so aufgebaut, daß durch geeignete Definitionen und nach den Beweisen von vier grundlegenden Sätzen eine vollständige formale Analogie zu der Darstellung in [18] gefunden wird. Dadurch ist es möglich, die für Experimente bedeutungsvollen Sätze ohne Beweise anzugeben. Die Beweise könnte man nämlich ganz genauso wie die in [18] durchführen und sie sind daher für das in dieser Arbeit interessierende Thema nicht mehr von Bedeutung.

Im ersten Abschnitt wird die allgemeine Theorie abgehandelt. Darauf folgend wird durch Einschränkungen der Wahlmöglichkeiten für die Kostenfunktion erreicht, schärfere Aussagen u.a. über Bayes-Experimentiervorschriften zu gewinnen. Zum Schluß dieses Kapitels wird dann am Beispiel der Vorgabeexperimente erläutert, wie man die bis zu diesem Zeitpunkt erhaltenen Aussagen verallgemeinern kann, wenn dem Experimentator nicht die volle Information über die bei dem Experiment mit dem Kanal ein- und ausgegebenen Worte zur Verfügung steht.

2.1 Allgemeine Theorie

2.1.1 Definitionen

K sei eine Menge von Kanälen über einem endlichen Eingabealphabet X^* und mit einem endlichen Ausgabealphabet Y . Ein Kanal aus K werde durch eine Funktion k mit

$$k: X^* \rightarrow \mathcal{M}^1(Y^*)$$

beschrieben. Dabei sei wieder $\mathcal{M}'(Y^*)$ die Menge aller Wahrscheinlichkeitsmaße über Y^* (die Bezeichnung ist sinnvoll, wenn man voraussetzt, daß die zugrunde liegende σ -Algebra über Y^* $\mathcal{P}(Y^*)$ ist). Der Einfachheit halber werde ein Kanal aus K selbst mit k bezeichnet. Es bezeichne weiter $k(\cdot|u)$ das Wk-maß, welches durch k an der Stelle $u \in X^*$ definiert ist. Jedes k aus K genügt den beiden folgenden Bedingungen:

$$(2.1) \left\{ \begin{array}{l} k(A|A) = 1 \\ \sum_{v \in Y^*: l(v)=l(u)} k(v|u) = 1 \end{array} \right. \quad \text{für alle } u \in X^*,$$

$$(2.2) \quad \sum_{v: l(v)=l(u)} k(v'v|u'u) = k(v'|u') \quad \text{für alle } u, u' \in X^* \\ \text{und alle } v' \in Y^* \text{ mit } l(v')=l(u').$$

Die Bedingung (2.2) sagt aus, daß die Kanäle aus K keinen Vorgriff besitzen (non anticipating), d.h. zukünftige Eingaben haben keinen Einfluß auf gegenwärtige Ausgaben. Die Kanäle aus K sollen sequentiell arbeiten, d.h. daß nach Eingabe eines $x \in X$ in einen Kanal genau ein $y \in Y$ vom Kanal ausgegeben wird und daß die Eingabe eines Wortes $u = x_1 x_2 \dots x_m$ bedeutet, daß nacheinander $x_1, x_2, \dots, x_m$ eingegeben werden.

A sei eine Menge von Aussagen über die Elemente von K und besitze die Elemente α als Aussagen.

Man stelle sich vor, genau $n \in \mathbb{N}$ Kopien eines Kanals $k \in K$ in die noch nichts eingegeben worden ist, liegen vor. In diese Kanäle können gleichzeitig Eingaben gemacht werden dergestalt, daß die Ausgabe einer Kopie auf eine Eingabe unabhängig von Eingaben und Ausgaben anderer Kopien ist. Weiter sei angenommen, dem Experimentator (ER) ist bekannt, daß $k \in K$, jedoch unbekannt, welches Element

es ist. Allein anhand von Eingaben in die einzelnen Kopien von k und den daraus resultierenden Ausgaben trifft der ER eine Aussage $\alpha \in A$ über den vorliegenden, dem ER aber in seiner Struktur unbekannten Kanal. Diese Vorgehensweise heiße n-fach Experiment (ET).

Ein ET wird charakterisiert durch eine Experimentiervorschrift (EV), gekennzeichnet durch eine statistische Entscheidungsfunktion ε

$$\varepsilon: \bigtimes_{i=1}^n \bigcup_{j=0}^{\infty} (X^j \times Y^j) \longrightarrow \mathcal{M}^1(A \cup \bigtimes_{i=1}^n X^* - \{(\Lambda, \dots, \Lambda)\}).$$

Es ist $X^j := \{u: u \in X^*, l(u) = j\}$ analog Y^j und

$\mathcal{M}^1(A \cup \bigtimes_{i=1}^n X^* - \{(\Lambda, \dots, \Lambda)\})$ die Menge aller Wk-maße über einer noch zu definierenden $\mathcal{G}$ -Algebra über der Argumentmenge. Mit $\varepsilon(\cdot | u_1, \dots, u_n, v_1, \dots, v_n)$ werde das durch ε an der Stelle $(u_1, v_1, \dots, u_n, v_n)$ mit $u_i \in X^*$ und $v_i \in Y^*$ festgelegte Wk-maß bezeichnet.

Ein ET, beschrieben durch eine EV ε , wird folgendermaßen durchgeführt:

Mit Hilfe des Wk-maßes $\varepsilon(\cdot | \Lambda, \dots, \Lambda, \Lambda, \dots, \Lambda)$ wird zufällig gewählt. Ist das Ergebnis ein $\alpha \in A$, so ist schon eine Aussage gefunden worden und man ist fertig. Im anderen Fall ist es ein $(u_1^1, \dots, u_n^1) \in \bigtimes_{i=1}^n X^*$. Dann gibt man das Wort u_i^1 in die i -te Kopie von k ($i=1, \dots, n$); die entsprechenden Reaktionen seien v_i^1 , auftretend mit den Wk $k(v_i^1 | u_i^1)$ ($i=1, \dots, n$).

Daraufhin wird mit $\varepsilon(\cdot | u_1^1, \dots, u_n^1, v_1^1, \dots, v_n^1)$ gewürfelt. Ist das Ergebnis ein $\alpha \in A$, so ist das Experiment beendet, andernfalls ist es ein $(u_1^2, \dots, u_n^2) \in \bigtimes_{i=1}^n X^*$. Dann wird u_i^2 in die i -te Kopie gegeben und man erhält v_i^2 ($i=1, \dots, n$). Die Gesamtausgabe-Wk sind dann $k(v_1^i v_2^i | u_1^i u_2^i)$ ($i=1, \dots, n$).

Anschließend wird mit $\varepsilon(\cdot | u_1^1 u_2^1, \dots, u_1^n u_2^n, v_1^1 v_2^1, \dots, v_1^n v_2^n)$ gewürfelt usw.

Ist bei der m -ten Entscheidung noch kein $a \in A$ gewürfelt worden, so ist es ein $(u_m^1, \dots, u_m^n)$ gewesen. Man gibt u_m^i in die i -te Kopie und erhält ein v_m^i mit der Gesamtausgabe-Wk $k(v_1^i \dots v_m^i | u_1^i \dots u_m^i)$ ($i = 1, \dots, n$). Dann wird mit $\varepsilon(\cdot | u_1^1 \dots u_m^1, \dots, u_1^n \dots u_m^n, v_1^1 \dots v_m^1, \dots, v_1^n \dots v_m^n)$ zufällig gewählt usw.

Diese Prozedur braucht nicht notwendig nach endlich vielen Entscheidungen abzubrechen.

Aus technischen Gründen werden folgende Schreibweisen eingeführt:

$$(2.3) \quad \left\{ \begin{array}{l} u_o = w_o = (\Lambda, \dots, \Lambda)_{n\text{-mal}}, \\ \text{für } m=1, 2, \dots \text{ mit } u_j^i \in X^* \\ u_m = \begin{pmatrix} u_1^1 & u_1^2 & \dots & u_1^n \\ u_2^1 & u_2^2 & \dots & u_2^n \\ \vdots & \vdots & & \vdots \\ u_m^1 & u_m^2 & \dots & u_m^n \end{pmatrix} \\ \text{und analog } w_m, \end{array} \right.$$

$$(2.4) \quad \varepsilon(\cdot | u_m, w_m) := \varepsilon(\cdot | u_1^1 \dots u_m^1, \dots, u_1^n \dots u_m^n, v_1^1 \dots v_m^1, \dots, v_1^n \dots v_m^n),$$

$$(2.5) \quad \ell(u_m) = (\ell(u_j^i))_{i=1, \dots, n; j=1, \dots, m},$$

$$(2.6) \quad \left\{ \begin{array}{l} k(w_o | u_o) = 1 \\ k(w_m | u_m) = \prod_{i=1}^n k(v_i^1 \dots v_i^m | u_i^1 \dots u_i^m). \end{array} \right.$$

Eine Funktion v mit den Eigenschaften:

$$v: K \times A \rightarrow \mathbb{R}^+ \cup \{0\},$$

$$v(k, a) \quad \left\{ \begin{array}{l} > 0, \text{ falls } a \text{ eine falsche Aussage über } k \text{ ist} \\ = 0 \text{ sonst} \end{array} \right.$$

heiße eine Verlustfunktion. Sie drückt im wesentlichen aus, wie schwerwiegend eine Fehlentscheidung ist. Wenn stets $v(k, a) \in \{0, 1\}$, dann ist v eine Fehlerfunktion.

Eine Kostenfunktion k ist eine nicht negative numerische Funktion, die für alle $m \in \mathbb{N}_0^{+}$ und alle v_m sowie w_m mit $\ell(v_m) = \ell(w_m)$ definiert ist. Sie soll zahlenmäßig den Aufwand beschreiben, den man benötigt, wenn v_m ein- und w_m ausgegeben⁺⁺ und anschließend eine Aussage aus A getroffen wird. Es sei $k(v_0, w_0) := 0$ mit der Begründung, daß a priori Entscheidungen keine Kosten verursachen.

Das mit einem ET zusammenhängende Entscheidungsproblem kann nun folgendermaßen formuliert werden:

n Kopien eines Kanales k aus einer bekannten Menge K liegen vor, jedoch ist unbekannt, welches Element von K es ist. Es soll eine Aussage aus A über den vorliegenden Kanal getroffen werden. Eine Verlust- und eine Kostenfunktion sind bekannt. Die Aufgabe besteht darin, eine EV ε für das ET zu finden.

Bevor Kriterien formuliert werden können, nach welchen eine EV gewählt werden könnte, sind weitere Definitionen erforderlich. Zunächst werden einige Wahrscheinlichkeiten aufgeschrieben.

$$(2.7) \quad p_1(v_m, A' | w_m, \varepsilon) := \prod_{i=0}^{m-1} \varepsilon(u_{i+1}^1, \dots, u_{i+1}^n | v_i, w_i) \cdot \varepsilon(A' | v_m, w_m)$$

drückt mit (2.3) und (2.4) die Wk dafür aus, daß v_m in die Kopien eingegeben und anschließend eine Aussage aus der (in noch festzulegendem Sinne meßbaren) Menge $A' \subset A$ gemacht wird, wenn von den Kopien w_m ausgegeben und die EV ε benutzt wird. Selbstverständlich ist

$$p_1(v_0, A' | w_0, \varepsilon) := \varepsilon(A' | v_0, w_0).$$

$$(2.8) \quad p_2(v_m, w_m, A' | k, \varepsilon) := p_1(v_m, A' | w_m, \varepsilon) k(w_m | v_m),$$

⁺) Sei $\mathbb{N}_0 := \{0, 1, 2, \dots\}$ im Gegensatz zu $\mathbb{N} = \{1, 2, \dots\}$.

⁺⁺) Diese vereinfachende Sprechweise besagt: für $i = 1, \dots, n$ wird in die i -te Kopie nacheinander $u_1^i, \dots, u_m^i$ eingegeben und daraufhin von dieser nacheinander $v_1^i, \dots, v_m^i$ ausgegeben.

wobei (2.6) angewendet wird, stellt die Wk dafür dar, daß v_m ein-, w_m ausgegeben und eine Aussage aus A' getroffen wird, wenn der Kanal k vorliegt und ε EV ist. Daß diese Interpretation sinnvoll ist, folgt aus der Eigenschaft der Kanäle aus K , keinen Vorgriff (vergl. (2.2)) zu besitzen. Mit (2.5) ist dann

$$(2.9) \quad p_3(v_m, A' | k, \varepsilon) := \sum_{w_m: \ell(w_m) = \ell(v_m)} p_2(v_m, w_m, A' | k, \varepsilon)$$

die Wk dafür, daß v_m eingegeben und eine Aussage aus A' getroffen wird, wenn k vorliegt und man ε benutzt. Schließlich ist

$$(2.10) \quad p_4(A' | k, \varepsilon) := \sum_{m=0}^{\infty} \sum_{v_m} p_3(v_m, A' | k, \varepsilon)$$

die Wk, daß man nach endlich vielen Entscheidungen eine Aussage aus A' trifft, wenn k vorliegt und ε benutzt wird. Offensichtlich ist es möglich, daß nicht immer $p_4(A | k, \varepsilon) = 1$ gilt.

Erfüllt die Verlustfunktion geeignete Meßbarkeitsbedingungen, so existiert

$$(2.11) \quad r_1(k, \varepsilon) := \int_A v(k, a) p_4(da | k, \varepsilon).$$

$r_1(k, \varepsilon)$ heißt bedingter erwarteter Verlust. Ist v eine Fehlerfunktion, dann ist $r_1(k, \varepsilon)$ der erwartete Entscheidungsfehler, wenn k vorliegt und ε benutzt wird. Mit der Vereinbarung $0 \cdot \infty = 0$ ($k(v_m, w_m)$ kann gleich ∞ sein) heißt

$$(2.12) \quad r_2(k, \varepsilon) := \sum_{m=1}^{\infty} \sum_{v_m} \sum_{w_m: \ell(v_m) = \ell(w_m)} k(v_m, w_m) p_2(v_m, w_m, A | k, \varepsilon)$$

bedingte erwartete Kosten.

$$(2.13) \quad r(k, \varepsilon) := r_1(k, \varepsilon) + r_2(k, \varepsilon)$$

wird dann bedingtes Risiko genannt.

Sei $\mathcal{L}_K$ eine geeignete σ -Algebra über K , die alle abzählbaren Teilmengen von K als Elemente enthält und $\bar{q}$ ein Wk-maß über $\mathcal{L}_K$. Ist $r(k, \varepsilon)$ meßbar bzgl. $\mathcal{L}_K$, so

existiert

$$(2.14) \quad \bar{r}(\bar{q}, \varepsilon) := \int_K r(k, \varepsilon) \bar{q}(dk)$$

mittleres Risiko genannt. $\bar{q}$ heit a priori Verteilung ber K .

Zum Schlu dieses Abschnittes werden noch einige besondere EV und a priori Verteilungen charakterisiert.

Eine EV ε' heit Bayes-EV bzgl $\bar{q}$, wenn

$$(2.15) \quad \bar{r}(\bar{q}, \varepsilon') \leq \bar{r}(\bar{q}, \varepsilon) \quad \text{fr alle } \varepsilon^+).$$

Eine EV ε heit abkrzend Bayes-EV, wenn es eine a priori Verteilung $\bar{q}$ gibt, so da ε Bayes-EV bzgl. $\bar{q}$ ist.

Eine EV ε' heit Minimax-EV, wenn

$$(2.16) \quad \sup_{k \in K} r(k, \varepsilon') \leq \sup_{k \in K} r(k, \varepsilon) \quad \text{fr alle } \varepsilon.$$

Eine a priori Verteilung $\bar{q}'$ heit am wenigsten gnstig, wenn

$$(2.17) \quad \inf_{\varepsilon} \bar{r}(\bar{q}', \varepsilon) \geq \inf_{\varepsilon} \bar{r}(\bar{q}, \varepsilon) \quad \text{fr alle a priori}$$

Verteilungen $\bar{q}$.

Eine EV ε_1 heit besser als eine EV ε_2 , wenn

$$(2.18) \quad \begin{cases} r(k, \varepsilon_1) \leq r(k, \varepsilon_2) & \text{fr alle } k \text{ und} \\ r(k, \varepsilon_1) < r(k, \varepsilon_2) & \text{fr mindestens ein } k. \end{cases}$$

Seien $\mathcal{E}_1$ und $\mathcal{E}_2$ mit $\mathcal{E}_1 \subset \mathcal{E}_2$ Mengen von EV. Dann heit $\mathcal{E}_1$ vollstndig relativ zu $\mathcal{E}_2$, wenn

$$(2.19) \quad \begin{cases} \text{fr alle } \varepsilon_2 \in \mathcal{E}_2 - \mathcal{E}_1, \text{ ein } \varepsilon_1 \in \mathcal{E}_1, \text{ existiert,} \\ \text{das besser als } \varepsilon_2 \text{ ist.} \end{cases}$$

⁺) d.h. fr alle ε aus der Menge $\mathcal{E}$ der dem ER zur Verfgung stehenden EV.

2.1.2 Voraussetzungen

Das Ziel des Kapitels 2 ist es, Aussagen über die in 2.1.1 genannten Begriffe und über deren Zusammenhänge zu gewinnen. Dieses Vorhaben ist nur zu verwirklichen, wenn einige Voraussetzungen gemacht werden. Dazu ist es notwendig, an einige topologische (vergl. z.B. [9]) und maßtheoretische (vergl. [3], [18]) Begriffe zu erinnern.

Zunächst wird eine Topologie in K definiert, die reguläre Topologie genannt wird, durch folgendes System offener Umgebungen der Elemente $k \in K$:

$$(2.20) \quad U(k) := \{ k' : |k(v_i | u_i) - k'(v_i | u_i)| < b \quad (i=1, \dots, t) \},$$

wobei $b \in \mathbb{R}^+$, $t \in \mathbb{N}$, $u_i \in X^*$, $v_i \in Y^*$ mit $l(v_i) = l(u_i)$ frei wählbar sind. Die reguläre Topologie ist offensichtlich hausdorffsch. Damit ist es sinnvoll, die reguläre Konvergenz einer Folge $(k_i)_{i \in \mathbb{N}}$ aus K gegen k_0 folgendermaßen festzulegen:

Für alle u und v mit $l(v) = l(u)$ gilt $\lim_{i \rightarrow \infty} k_i(v|u) = k_0(v|u)$.

K2.1 Korollar: K ist separabel bzgl. der regulären Topologie.

Beweis: (Es ist zu zeigen, daß es eine Folge $(k_i)_{i \in \mathbb{N}}$ aus K gibt mit der Eigenschaft, daß zu jedem $k \in K$ eine Teilfolge von $(k_i)_{i \in \mathbb{N}}$ existiert, die regulär gegen k konvergiert.) Offensichtlich gibt es für jedes $u \in X^*$ eine Folge $(k_{u_i})_{i \in \mathbb{N}}$ aus K , die bzgl. der Pseudometrik in K

$$(2.21) \quad d_u(k_1, k_2) := \max_{v: l(v)=l(u)} |k_1(v|u) - k_2(v|u)|$$

dicht in K liegt. Für alle $k \in K$ und eine Folge $(b_u)_{u \in X^*}$ (bei geeigneter Abzählung von X^*) positiver reeller Zahlen mit $\lim_u b_u = 0$ gilt: für alle $u \in X^*$ gibt es ein $i_u \in \mathbb{N}$ mit $d_u(k, k_{u_{i_u}}) \leq b_u$. $d_u(k_1, k_2)$ ist mit "wachsendem" u nicht fallend, demzufolge konvergiert $(k_{u_{i_u}})_{u \in X^*}$ regulär gegen k .]

Für alle $k \in K$ gibt es eine abzählbare offene Umgebungsbasis (man wähle z.B. b rational in (2.20)). Mit K2.1 existiert dann eine abzählbare Basis offener Mengen in K . K ist aber auch lokal-kompakt, demnach mit [3] S.44.1 polnisch. D.h. K besitzt eine die reguläre Topologie definierende Metrik, und der dadurch bestimmte metrische Raum ist vollständig.

Nachdem die Struktur von K festgelegt worden ist, soll diejenige der Aussagenmenge A festgelegt werden. Dazu die ersten Forderungen:

V2.1 a) $v(k, a)$ soll eine beschränkte Funktion von k und a sein.

b) Sei für $a_1, a_2 \in A$

$$d_A(a_1, a_2) := \sup_{k \in K} |v(k, a_1) - v(k, a_2)|.$$

Für alle $k \in K$ sei $v(k, a)$ eine bezüglich der durch d_A in A definierten Topologie stetige Funktion in A .

V2.2 a) d_A sei Metrik (genannt eigentliche Metrik) in A .

b) A sei kompakt im Sinne der eigentlichen Metrik d_A .

Sei

$$E := A \cup \bigcup_{i=1}^{\infty} X^* - \{u_0\}.$$

Da A metrisch und kompakt bzgl. d_A ist, besitzt A eine abzählbare Basis C offener Mengen in A . Die Menge der bzgl. d_A offenen Teilmengen von A vereinigt mit der Potenzmenge von $\bigcup_{i=1}^{\infty} X^* - \{u_0\}$ definiert eine Topologie offener Mengen in E . $C \cup [\bigcup_{i=1}^{\infty} X^* - \{u_0\}]$ ist dann eine abzählbare Basis in E . E ist hausdorffsch und lokalkompakt, da A kompakt ist, demnach auch polnisch, und es gibt eine seine Topologie definierende vollständige Metrik. E stellt eine K_b -Menge dar, demnach ist E im Unendlichen abzählbar.

Sei $\mathfrak{X}_E$ die kleinste $\mathcal{C}$ -Algebra, die alle im eben erläuterten Sinne offenen Teilmengen von E enthält und $\mathcal{M}^c(E)$ die Menge aller endlichen Baire-Maße μ über $\mathfrak{X}_E$. Wenn $\mathcal{C}^b(E)$ die Menge aller auf E beschränkten und stetigen Funktionen ist, dann heißt die größte aller Topologien, bzgl. der alle Funktionen der Art

$$\mu \rightarrow \int_E f d\mu \quad (f \in \mathcal{C}^b(E))$$

stetig sind, schwache Topologie in $\mathcal{M}^c(E)$. Diese wird erzeugt durch die folgenden offenen Umgebungen eines beliebigen Elementes μ_0 von $\mathcal{M}^c(E)$

$$U(\mu_0) := \{ \mu : | \int_E f_i d\mu - \int_E f_i d\mu_0 | < b \quad (i=1, \dots, t) \},$$

wobei $b \in \mathbb{R}^+$, $t \in \mathbb{N}$, $f_i \in \mathcal{C}^b(E)$ frei wählbar sind.

Die schwache Topologie ist hausdorffsch nach dem Rieszschen Darstellungssatz ([3] S. 43.4); sie ist sogar metrisierbar, da E eine abzählbare Basis besitzt ([3] S. 46.9).

Eine Folge $(\mu_i)_{i \in \mathbb{N}_0}$ aus $\mathcal{M}^c(E)$ heißt dann schwach konvergent gegen μ_0 , wenn gilt

$$(2.22) \quad \int_E f d\mu_i \xrightarrow{i \rightarrow \infty} \int_E f d\mu_0 \quad \text{für alle } f \in \mathcal{C}^b(E).$$

Eine Menge $B \subset \mathcal{M}^c(E)$ heißt beschränkt, wenn

$$\sup_{\mu \in B} \mu(E) < \infty.$$

B ist genau dann schwach relativ kompakt, wenn B beschränkt ist. Ist B zudem noch schwach abgeschlossen, so ist B schwach kompakt. B ist schließlich genau dann schwach kompakt, wenn jede Folge in B eine schwach konvergente Teilfolge besitzt.

Mit den Bemerkungen des letzten Absatzes ist man nun in der Lage, einen sinnvollen Konvergenzbegriff in einer Menge $\mathcal{E}$ von EV zu definieren. Eine Folge $(\varepsilon_i)_{i \in \mathbb{N}_0}$ von EV aus $\mathcal{E}$ heißt regulär konvergent in $\mathcal{E}$ gegen ε_0 , wenn für alle m , alle v_m und alle w_m mit $\ell(w_m) = \ell(v_m)$ $(\varepsilon_i(\cdot | v_m, w_m))_{i \in \mathbb{N}_0}$ schwach (vergl. (2.22)) gegen $\varepsilon_0(\cdot | v_m, w_m)$ konvergiert. Es ist klar, was dann unter

der regulären Topologie in $\mathcal{E}$ zu verstehen ist und braucht nicht mehr aufgeschrieben zu werden.

Die Menge $\mathcal{E}$ aller dem ER zur Verfügung stehenden EV wird durch die folgenden Forderungen eingeschränkt:

- V2.3 a) Für jedes $s \in \mathbb{N}_0$ gibt es ein $b_s \in \mathbb{N}_0$, das nur von s abhängt, derart, daß für alle $m \in \mathbb{N}_0$ und $v_m = (u_j^i)$ mit $\max_{i=1, \dots, n} l(u_1^i \dots u_m^i) = s^+$, alle w_m mit $\ell(w_m) = \ell(v_m)$ sowie alle $\varepsilon \in \mathcal{E}$ gilt:
 Falls $\max_{i=1, \dots, n} l(u_{m+1}^i) > b_s$, dann
 $\varepsilon(u_{m+1}^1, \dots, u_{m+1}^n | v_m, w_m) = 0$.

b) Sei für $m \in \mathbb{N}$

$$(2.23) \quad p_s(v_m | w_{m-1}, \varepsilon) := \prod_{i=0}^{m-1} \varepsilon(u_{i+1}^1, \dots, u_{i+1}^n | v_i, w_i)$$

die Wk dafür, daß v_m gewürfelt, wenn w_{m-1} vom Kanal ausgegeben und ε benutzt wird. Falls für ein festes v_m die Kostenfunktion k die Eigenschaft hat $k(v_m, w_m) = \infty$ für alle w_m mit $\ell(w_m) = \ell(v_m)$, dann sei für alle w_{m-1} und $\varepsilon \in \mathcal{E}$
 $p_s(v_m | w_{m-1}, \varepsilon) = 0$.

Nach V2.3 a) ist es also bei einer Entscheidung f.s. nicht möglich, beliebig lange Eingabewörter für die Kopien zu wählen. Die Forderung V2.3 b) zeigt, daß $\mathcal{E}$ von der gewählten Kostenfunktion k abhängt. Eine Kostenfunktion möge selbst den folgenden Voraussetzungen genügen:

V2.4 a) Für $m \in \mathbb{N}_0$ gelte stets

$$0 \leq k(v_m, w_m) \leq k(v_{m+1}, w_{m+1}).$$

b) Es gibt eine Folge $(b_s')_{s \in \mathbb{N}_0}$ mit $\lim_{s \rightarrow \infty} b_s' = \infty$, so daß gilt:

$$\text{Wenn } \max_{i=1, \dots, n} l(u_1^i \dots u_m^i) \geq s, \text{ dann} \\ k(v_m, \cdot) \geq b_s'.$$

⁺) falls $m=0$ ist, sei $l(u_1 \dots u_m) := 0$.

V2.4 c) Für ein festes v_m ist $k(v_m, \cdot)$ entweder beschränkt oder identisch ∞ .

Mit V2.4 a), b) erfüllen die Kostenfunktionen die sinnvolle Forderung, daß längere Experimente mehr kosten als kürzere und daß die Kosten mit der Länge des Experimentes gegen unendlich gehen. V2.4 c) ist im Zusammenhang mit V2.3 b) zu sehen und dient zur Festlegung der Menge $\mathcal{E}$ der EV.

Eine Menge $\mathcal{E}'$ von EV heißt konvex, wenn für je zwei $\varepsilon_1, \varepsilon_2 \in \mathcal{E}'$ und jedes $\alpha \in]0, 1[$ ein $\varepsilon \in \mathcal{E}'$ existiert, so daß mit (2.23) und (2.7) stets

$$a) \quad p_S(v_m | \omega_{m-1}, \varepsilon) = \alpha p_S(v_m | \omega_{m-1}, \varepsilon_1) + (1-\alpha) p_S(v_m | \omega_{m-1}, \varepsilon_2),$$

$$b) \quad \varepsilon(A' | v_0, \omega_0) = \alpha \varepsilon_1(A' | v_0, \omega_0) + (1-\alpha) \varepsilon_2(A' | v_0, \omega_0),$$

$$\text{wenn } A' \in \mathcal{X}_A := \{B \cap A \mid B \in \mathcal{X}_\varepsilon\},$$

$$c) \quad p_1(v_m, A' | \omega_m, \varepsilon) = \alpha p_1(v_m, A' | \omega_m, \varepsilon_1) + (1-\alpha) p_1(v_m, A' | \omega_m, \varepsilon_2),$$

wenn $A' \in \mathcal{X}_A$.

Nun sind sofort drei Eigenschaften der Menge $\mathcal{E}$, der dem ER zur Verfügung stehenden EV, die nach Definition einer V2.4 genügenden Kostenfunktion und wegen V2.3 festgelegt ist, angebar. Die Beweise können durch einfaches Nachprüfen der zu zitierenden Eigenschaften unter Verwendung von V2.3 erbracht werden.

K2.2 Korollar: $\mathcal{E}$ ist konvex.

K2.3 Korollar: $\mathcal{E}$ ist regulär abgeschlossen.

K2.4 Korollar: Eine EV ε gehört zu $\mathcal{E}$, wenn es eine EV $\varepsilon' \in \mathcal{E}$ und ein $\alpha' \in A$ derart gibt, daß für jedes v_m und jedes ω_m mit $\ell(\omega_m) = \ell(v_m)$ entweder

$$\varepsilon(\cdot | v_m, \omega_m) = \varepsilon'(\cdot | v_m, \omega_m) \quad \text{oder}$$

$$\varepsilon(\alpha' | v_m, \omega_m) = 1.$$

Zum Schluß dieses Abschnittes muß noch auf einige Meßbarkeitseigenschaften eingegangen werden.

Wenn B eine abzählbare Menge ist, dann ist die Potenzmenge von B die kleinste σ -Algebra über B , die alle Einermengen enthält. Deshalb soll für das Folgende stets angenommen sein, daß die Potenzmenge einer abzählbaren Menge die zugrunde liegende σ -Algebra ist. Meßbarkeitsfragen spielen in diesem Zusammenhang dann keine Rolle mehr.

Als σ -Algebra $\mathcal{X}_K$ über K wird nun die kleinste σ -Algebra benutzt, die alle regulär offenen Teilmengen von K enthält. In A wird eine σ -Algebra in Einklang mit $\mathcal{X}_E$ als kleinste σ -Algebra festgelegt, die alle bzgl. der Metrik d_A offenen Teilmengen von A enthält. $\mathcal{X}_{K \times A}$ als kleinste σ -Algebra, die $\mathcal{X}_K \times \mathcal{X}_A$ enthält, sei dann die interessierende σ -Algebra auf $K \times A$.

V2.5 a) Die Verlustfunktion $v(k, a)$ sei meßbar bzgl. $\mathcal{X}_{K \times A}$.

b) $f(v|u, k) := k(v|u)$ sei meßbar bzgl. $\mathcal{X}_K$ für alle $u \in X^*$ und $v \in Y^*$ mit $l(v) = l(u)$.

Nun läßt sich zeigen, daß die in (2.13) und (2.14) definierten Funktion r und $\bar{r}$ existieren.

Aus V2.5 b) folgt, daß für alle $A' \in \mathcal{X}_A$ und alle $\varepsilon \in \mathcal{E}$ mit (2.10) $p_4(A'|k, \varepsilon)$ meßbar bzgl. $\mathcal{X}_K$ ist. A ist kompakt bzgl. d_A und v beschränkt sowie meßbar bzgl. $\mathcal{X}_{K \times A}$ nach V2.5 a). Dann existiert mit (2.11)

$$r_1(k, \varepsilon) = \int_A v(k, a) p_4(da | k, \varepsilon) < \infty.$$

$r_1(k, \varepsilon)$ kann als Grenzwert einer Folge

$$r_1^i(k, \varepsilon) := \sum_{j=1}^{i} v(k, a_j^i) p_4(A_j^i | k, \varepsilon) \quad (i \in \mathbb{N})$$

mit $A_j^i \in \mathcal{L}_A, a_j^i \in A$ und $l_i \in \mathbb{N}$ monoton wachsend aufgefaßt werden. $r_1^i(k, \varepsilon)$ ist meßbar bzgl. $\mathcal{L}_K$, also auch $r_1(k, \varepsilon) \cdot r_2(k, \varepsilon)$ existiert wegen (2.12) und (2.8), da R eine numerische nichtnegative Funktion ist und ist aus analogen Gründen wie $r_1(k, \varepsilon)$ meßbar bzgl. $\mathcal{L}_K$. Dann ist es aber auch $r(k, \varepsilon) = r_1(k, \varepsilon) + r_2(k, \varepsilon)$. $\bar{q}$ ist ein Maß über $\mathcal{L}_K$, dann existiert schließlich $\bar{r}(\bar{q}, \varepsilon) = \int_K r(k, \varepsilon) \bar{q}(dk)$.

2.1.3 Aussagen

Im letzten Abschnitt sind die Voraussetzungen geschaffen worden, um die Eigenschaften, die in 2.1.1 definiert worden sind, näher beleuchten zu können. Dazu wird im folgenden stets angenommen, daß das Entscheidungsproblem festgelegt und die Voraussetzungen V2.1 bis V2.5 immer erfüllt seien. Die zunächst zu beweisenden Sätze befassen sich mit einer Kompaktheitseigenschaft von $\mathcal{E}$, mit Stetigkeitsfragen von $\bar{r}$ und der Separabilität von K bzgl. einer von r abhängigen Pseudometrik. Dann ist es gelungen, die hier behandelte Theorie zu einer vollständigen Analogie mit der Theorie der statistischen Entscheidungsfunktionen nach [18] zu bringen.

S2.5 Satz: Jede Folge $(\varepsilon_i)_{i \in \mathbb{N}}$ aus $\mathcal{E}$ enthält eine in $\mathcal{E}$ regulär konvergente Teilfolge.

Beweis: Sei $s \in \mathbb{N}_0$ und b_s die in V2.3 a) festgelegte Zahl in Abhängigkeit von s . Mit $E_s \subset E$ werde die Vereinigung von A und der Menge aller $(u_1^i, \dots, u_n^i) \in \prod_{i=1}^n X^*$ mit $\max_{i=1, \dots, n} l(u^i) \leq b_s$ bezeichnet. Dann ist nach V2.3 a) für alle $s \in \mathbb{N}_0$, alle $m \in \mathbb{N}_0$, alle v_m mit $s = \max_{i=1, \dots, n} l(u_1^i \dots u_m^i)$, alle w_m mit

$\ell(\mu_m) = \ell(\nu_m)$ und alle $\varepsilon \in \mathcal{E}$

$\varepsilon(E_s | \nu_m, \mu_m) = 1$.

Wegen V2.2 b) ist A bzgl. d_A kompakt und E_s enthält außer den Elementen von A nur endlich viele Punkte. Daher ist E_s kompakt für jedes $s \in \mathbb{N}_0$.

Seien ν_m und μ_m mit $\max_{i=1, \dots, n} \sum_{j=1}^m l(u_j^i) = s$ und $\ell(\mu_m) = \ell(\nu_m)$

fest gewählt. Dann kann die Folge $(\varepsilon_i(\cdot | \nu_m, \mu_m))_{i \in \mathbb{N}}$ als eine Folge von Wk-maßen über E_s (mit der entsprechenden Spur der ε -Algebra $\mathcal{L}_\varepsilon$) aufgefaßt werden. E_s ist kompakt und damit ist $\mathcal{M}^1(E_s)$ schwach abgeschlossen.

Mithin gehören alle Häufungspunkte der Menge

$\{\varepsilon_i(\cdot | \nu_m, \mu_m) : i \in \mathbb{N}\}$ zu $\mathcal{M}^1(E_s)$. Dann enthält die Folge $(\varepsilon_i(\cdot | \nu_m, \mu_m))_{i \in \mathbb{N}}$ eine gegen ein $\varepsilon_0(\cdot | \nu_m, \mu_m) \in \mathcal{M}^1(E_s)$ konvergente Teilfolge.

Da die Menge aller Argumente einer EV ε abzählbar ist, läßt sich mit dem Cantorschen Verfahren aus $(\varepsilon_i)_{i \in \mathbb{N}}$ eine Teilfolge gewinnen, die für alle Argumente schwach konvergiert; also konvergiert die Teilfolge auch regulär, da mit K2.3 der Grenzwert der Teilfolge zu $\mathcal{E}$ gehört.]

S2.6 Satz: Die Folge $(\varepsilon_i)_{i \in \mathbb{N}_0}$ aus $\mathcal{E}$ konvergiere regulär gegen ε_0 , dann ist für jedes Wk-maß $\bar{q}$ über $\mathcal{L}_\mu$

$$\lim_{i \rightarrow \infty} \bar{r}(\bar{q}, \varepsilon_i) \geq \bar{r}(\bar{q}, \varepsilon_0).$$

Zu jeder Folge $(\varepsilon_i)_{i \in \mathbb{N}}$ aus $\mathcal{E}$ existiert eine Teilfolge $(\varepsilon_{i_j})_{j \in \mathbb{N}}$ und ein $\varepsilon_0 \in \mathcal{E}$, so daß für alle $\bar{q}$ gilt

$$\lim_{j \rightarrow \infty} \bar{r}(\bar{q}, \varepsilon_{i_j}) \geq \bar{r}(\bar{q}, \varepsilon_0).$$

Beweis: Der zweite Teil der Behauptung folgt unmittelbar aus dem ersten Teil und S2.5. Es braucht also nur der erste Teil bewiesen zu werden.

Die Folge $(\varepsilon_i)_{i \in \mathbb{N}}$ konvergiere regulär gegen ε_0 in $\mathcal{E}$. Falls für ein $\bar{q}$ $\lim_{i \rightarrow \infty} \bar{r}(\bar{q}, \varepsilon_i) = \infty$, ist nichts zu beweisen. Sei also für ein $\bar{q}$ $\lim_{i \rightarrow \infty} \bar{r}(\bar{q}, \varepsilon_i) < \infty$.

Wegen der regulären Konvergenz von $(\varepsilon_i)_{i \in \mathbb{N}}$ konvergiert für jedes $m \geq 0$, v_m, w_m mit $\ell(w_m) = \ell(v_m)$ und jedes $(u^1, \dots, u^n) \in (\varepsilon_i(u^1, \dots, u^n | v_m, w_m))_{i \in \mathbb{N}}$ gegen $\varepsilon_0(u^1, \dots, u^n | v_m, w_m)$, da die Indikatorfunktion von $(u^1, \dots, u^n)$ aus $\mathcal{C}^b(E)$ und $(\varepsilon_i(\cdot | v_m, w_m))_{i \in \mathbb{N}}$ schwach gegen $\varepsilon_0(\cdot | v_m, w_m)$ konvergiert.

Dann folgt mit (2.7), (2.8) und (2.9) für jedes $f \in \mathcal{C}^b(E)$ und $k \in K$

$$(2.24) \quad \lim_{i \rightarrow \infty} \int_A f(a) \rho_2(v_m, w_m, da | k, \varepsilon_i) = \int_A f(a) \rho_2(v_m, w_m, da | k, \varepsilon_0),$$

$$(2.25) \quad \lim_{i \rightarrow \infty} \int_A f(a) \rho_3(v_m, da | k, \varepsilon_i) = \int_A f(a) \rho_3(v_m, da | k, \varepsilon_0).$$

Die Integrale in (2.24) und (2.25) sind als Funktionen von k bzgl. $\mathcal{B}_K$ meßbar und sind beschränkt, da f beschränkt ist. Dann gilt mit dem Satz von der majorisierten Konvergenz von Lebesgue:

$$(2.26) \quad \lim_{i \rightarrow \infty} \int_K \int_A f(a) \rho_2(v_m, w_m, da | k, \varepsilon_i) \bar{q}(dk) = \int_K \int_A f(a) \rho_2(v_m, w_m, da | k, \varepsilon_0) \bar{q}(dk),$$

$$(2.27) \quad \lim_{i \rightarrow \infty} \int_K \int_A f(a) \rho_3(v_m, da | k, \varepsilon_i) \bar{q}(dk) = \int_K \int_A f(a) \rho_3(v_m, da | k, \varepsilon_0) \bar{q}(dk).$$

Dann folgt aus (2.26), daß

$$(2.28) \quad \rho_6(v_m, w_m, A | \bar{q}, \varepsilon_i) := \int_K \rho_2(v_m, w_m, A | k, \varepsilon_i) \bar{q}(dk)$$

mit $i \rightarrow \infty$ gegen $\rho_6(v_m, w_m, A | \bar{q}, \varepsilon_0)$ konvergiert.

v ist eine bzgl. $\mathcal{L}_{K \times A}$ meßbare, beschränkte und eine stetige Funktion von a , daher existiert

$$(2.29) \quad \bar{r}_1(\bar{q}, \varepsilon_i, v_m) := \int_K \int_A v(k, a) p_3(v_m, da | k, \varepsilon_i) \bar{q}(dk)$$

und konvergiert für $i \rightarrow \infty$ mit (2.27) gegen $\bar{r}_1(\bar{q}, \varepsilon_0, v_m)$.

Aus V2.4 c) und V2.3 b) sowie (2.23), (2.7) und (2.8) folgt: wenn $k(v_m, w_m) = \infty$, dann ist $p_2(v_m, w_m, A | k, \varepsilon) = 0$.

Daher konvergiert mit der Vereinbarung $\infty \cdot 0 = 0$ und (2.28)

$$(2.30) \quad \bar{r}_2(\bar{q}, \varepsilon_i, v_m) := \sum_{w_m: k(v_m, w_m) = \ell(v_m)} k(v_m, w_m) p_6(v_m, w_m, A | \bar{q}, \varepsilon_i)$$

mit $i \rightarrow \infty$ gegen $\bar{r}_2(\bar{q}, \varepsilon_0, v_m)$.

Sei mit (2.11) und (2.12) für $i \in \mathbb{N}_0$

$$(2.31) \quad \bar{r}_j(\bar{q}, \varepsilon_i) := \int_K r_j(k, \varepsilon_i) \bar{q}(dk) \quad (j=1,2),$$

dann ist

$$\bar{r}(\bar{q}, \varepsilon_i) = \bar{r}_1(\bar{q}, \varepsilon_i) + \bar{r}_2(\bar{q}, \varepsilon_i).$$

Es gilt mit (2.11), (2.29) und (2.31)

$$\bar{r}_1(\bar{q}, \varepsilon_i) = \sum_{m=0}^{\infty} \sum_{v_m} \bar{r}_1(\bar{q}, \varepsilon_i, v_m)$$

aus folgendem Grund:

$$\mu_t := \sum_{m=0}^t \sum_{v_m} p_3(v_m, \cdot | k, \varepsilon_i)$$

ist für $t \in \mathbb{N}_0$ ein Maß über $\mathcal{G}_A$ aus der Menge

$\mathcal{M}'(A) := \{ \mu : \mu(A) \leq 1 \}$. $\mathcal{M}'(A)$ ist schwach kompakt,

d.h. $(\mu_t)_{t \in \mathbb{N}_0}$ besitzt eine gegen ein $\mu \in \mathcal{M}'(A)$

schwach konvergente Teilfolge $(\mu_{t_i})_{i \in \mathbb{N}}$. Für jedes

$A' \in \mathcal{G}_A$ gilt $\mu_t(A') \leq \mu_{t+1}(A')$, demzufolge konver-

giert $(\mu_t)_{t \in \mathbb{N}_0}$ selbst schwach gegen μ . Somit gilt

nach Definition der schwachen Konvergenz und wegen der

Stetigkeit von $v(k, a)$ bzgl. a die Vertauschbarkeit

von $\int_A$ und $\sum_m \sum_{v_m}$. $\int_K$ und $\sum_m \sum_{v_m}$ können wegen

der Beschränktheit von v nach dem Satz von der majorisier-

ten Konvergenz vertauscht werden.

Es gilt weiterhin mit (2.12), (2.30) und (2.31) für $i \in \mathbb{N}_0$

$$\bar{r}_2(\bar{q}, \varepsilon_i) = \sum_{m=0}^{\infty} \sum_{u_m} \bar{r}_2(\bar{q}, \varepsilon_i, u_m),$$

da R eine nicht negative numerische Funktion ist und man den Satz von der monotonen Konvergenz von Levi anwenden kann.

Sei nun für $i, s \in \mathbb{N}_0$

$$\bar{r}^s(\bar{q}, \varepsilon_i) := \sum_{t=1}^2 \sum_{m=0}^s \sum_{u_m} \bar{r}_t(\bar{q}, \varepsilon_i, u_m).$$

Dann ist für alle $i \in \mathbb{N}_0$ die Folge $(\bar{r}^s(\bar{q}, \varepsilon_i))_{s \in \mathbb{N}_0}$ monoton wachsend und gegen $\bar{r}(\bar{q}, \varepsilon_i)$ konvergierend.

Andererseits konvergiert $(\bar{r}^s(\bar{q}, \varepsilon_i))_{i \in \mathbb{N}}$ gegen $\bar{r}^s(\bar{q}, \varepsilon_0)$ für alle $s \in \mathbb{N}_0$. Daraus kann man elementar, z.B. indirekt, auf die erste Behauptung des Satzes schließen.]

K2.7 Korollar: Sei $(\varepsilon_i)_{i \in \mathbb{N}_0}$ eine Folge von EV, für die die Wk eins ist, daß das ET nach höchstens m Entscheidungen ausgeführt ist, d.h. ein Element aus A gewählt ist, und die regulär gegen ε_0 konvergiert. Dann gilt für alle a priori Wk-maße $\bar{q}$

$$\lim_{i \rightarrow \infty} \bar{r}(\bar{q}, \varepsilon_i) = \bar{r}(\bar{q}, \varepsilon_0).$$

Dieser Sachverhalt folgt unmittelbar aus dem Beweis von S2.6.

Sei $\mathcal{E}^m$ für ein $m \in \mathbb{N}$ die Menge aller EV ε^m , die ET induzieren, welche mit der Wk 1 nach spätestens m Entscheidungen ausgeführt sind. Dann gibt es mit V2.3 a) eine natürliche Zahl $l' = l'(m)$, die nur von m abhängt, so daß l' das Maximum aller Ausdrücke $\max_{i=1, \dots, n} l(u_1^i \dots u_{m-1}^i)$

über alle u_{m-1} unter der Einschränkung, daß $u_{m-1} = (u_j^i)$ von irgendeinem $\varepsilon \in \mathcal{E}$ mit positiver Wk gewählt werden kann, darstellt.

Zur Erläuterung des Problems im nächsten Satz werden im folgenden Absatz vier Pseudometriken in K definiert:

$$(2.32) \quad d_1^m(k_1, k_2) := \max_{u: l(u) \leq l'} d_u(k_1, k_2) = \max_{u: l(u) \leq l'} \max_{v: l(v) = l(u)} |k_1(v|u) - k_2(v|u)|,$$

wobei d_u in (2.21) definiert wurde,

$$(2.33) \quad d_2^m(k_1, k_2) := \sup_{\varepsilon^m \in \mathcal{E}^m} |r(k_1, \varepsilon^m) - r(k_2, \varepsilon^m)|,$$

$$(2.34) \quad d_3^m(k_1, k_2) := \sup_{a \in A} |v(k_1, a) - v(k_2, a)|,$$

$$(2.35) \quad d_4^m(k_1, k_2) := d_1^m + d_3^m.$$

Die d_i^m definieren in K Umgebungstopologien, die i.a. nicht hausdorffsch sind. K heißt nun separabel bzgl. d_i^m , wenn es eine abzählbare bzgl. d_i^m in K dichte Teilmenge von K gibt.

S2.8 Satz: K ist separabel bzgl. d_2^m für jedes $m \in \mathbb{N}$.

Beweis: Aus K2.1 folgt, daß K separabel bzgl. d_1^m (vergl. (2.32)) ist. Es wird zunächst gezeigt, daß K bzgl. d_4^m (vergl. (2.35)) separabel ist, und daraus läßt sich dann die Behauptung folgern.

A ist wegen V2.2 b) kompakt bzgl. der Metrik d_A , also total beschränkt. D.h. zu jedem $b \in \mathbb{R}^+$ gibt es eine endliche Menge von Punkten $A(b) := \{a_1, \dots, a_s\} \subset A$, so daß für alle $a \in A$ ein $a_i \in A(b)$ existiert mit

$$d_A(a, a_i) = \sup_{k \in K} |v(k, a) - v(k, a_i)| < b.$$

Sei über $K \times K$ definiert

$$d_{3A(b)}^m(k_1, k_2) := \max_{a_i \in A(b)} |v(k_1, a_i) - v(k_2, a_i)|.$$

Da mit V2.1 a) $v(k, a)$ beschränkt ist, gibt es zu jedem $b \in \mathbb{R}^+$ eine endliche Teilmenge $K(b) \subset K$, so daß für alle $k \in K$ ein $k_i \in K(b)$ existiert mit der Eigenschaft $d_{3A(b)}^m(k, k_i) < b$.

Diese Behauptung ist aus folgendem Grund richtig: Für jede Folge $(k_i)_{i \in \mathbb{N}}$ aus K und jedes $a \in A$ enthält die Zahlenfolge $(v(k_i, a))_{i \in \mathbb{N}}$ eine konvergente Teilfolge. Aus $(k_i)_{i \in \mathbb{N}}$ läßt sich dann eine Teilfolge $(k_{i_j})_{j \in \mathbb{N}}$ auswählen, die bzgl. $d_{3A(b)}^m$ eine Cauchyfolge ist. Damit läßt sich die totale Beschränktheit von K bzgl. $d_{3A(b)}^m$ beweisen. Es gilt mit (2.34)

$$\begin{aligned} & |d_3^m(k_1, k_2) - d_{3A(b)}^m(k_1, k_2)| \\ &= | \max_{a \in A} |v(k_1, a) - v(k_2, a)| - \max_{a_i \in A(b)} |v(k_1, a_i) - v(k_2, a_i)| | \\ &\leq ||v(k_1, a_0) - v(k_2, a_0)| - |v(k_1, a_{i_0}) - v(k_2, a_{i_0})||, \end{aligned}$$

wobei a_0 Extrempunkt von $|v(k_1, a) - v(k_2, a)|$ und $a_{i_0} \in A(b)$ so gewählt ist, daß $d_A(a_0, a_{i_0}) < b$ ist. Dann schätzt man weiter ab

$$\begin{aligned} & |d_3^m(k_1, k_2) - d_{3A(b)}^m(k_1, k_2)| \\ &\leq |v(k_1, a_0) - v(k_1, a_{i_0})| + |v(k_2, a_0) - v(k_2, a_{i_0})| \leq 2b. \end{aligned}$$

Schließlich gibt es zu jedem $k \in K$ ein $k_i \in K(b)$, so daß

$$d_3^m(k, k_i) \leq d_{3A(b)}^m(k, k_i) + |d_3^m(k, k_i) - d_{3A(b)}^m(k, k_i)| \leq 3b.$$

Somit ist K total beschränkt bzgl. d_3^m .

Sei $b \in \mathbb{R}^+$, dann gibt es ein $s \in \mathbb{N}$ und eine direkte Zerlegung $\{K_i : i=1, \dots, s\}$ von K mit der Eigenschaft

$$\sup_{k_1, k_2 \in K_i} d_j^m(k_1, k_2) < b \quad (i=1, \dots, s).$$

K ist separabel bzgl. d_1^m . Demnach gibt es abzählbare Mengen K_i' aus K_i , die dicht bzgl. d_1^m in K_i liegen ($i=1, \dots, s$). Dann ist $K' := \bigcup_{i=1}^s K_i'$ abzählbar und hat die Eigenschaft:

Für jedes $k \in K$ gibt es genau ein K_i mit $k \in K_i$ und ein $k_1 \in K_i' \subset K'$ mit (vergl. (2.35))

$$d_4^m(k, k_1) = d_1^m(k, k_1) + d_3^m(k_1, k_1) < 2b.$$

Also ist K separabel bzgl. d_4^m .

Der Satz ist bewiesen, wenn gezeigt werden kann:

Gilt für eine Folge $(k_i)_{i \in \mathbb{N}_0}$ aus K

$$(2.36) \quad \lim_{i \rightarrow \infty} d_4^m(k_i, k_0) = 0,$$

so ist auch mit (2.33)

$$\lim_{i \rightarrow \infty} d_2^m(k_i, k_0) = 0.$$

Sei nun (2.36) gültig für eine Folge $(k_i)_{i \in \mathbb{N}_0}$, dann folgt $\lim_{i \rightarrow \infty} v(k_i, a) = v(k_0, a)$ gleichmäßig in a und

$$\lim_{i \rightarrow \infty} k_i(v|u) = k_0(v|u) \quad \text{gleichmäßig in } v, u$$

mit $l(v) = l(u) \leq l'(m)$.

Da mit (2.13) offensichtlich

$$r(k, \varepsilon^m) = \sum_{s=0}^{m-1} \sum_{v_s, \omega_s} \left[\int_A v(k, a) k(\omega_s | v_s) p_1(v_s, da | \omega_s, \varepsilon^m) \right. \\ \left. + k(v_s, \omega_s) k(\omega_s | v_s) p_1(v_s, A | \omega_s, \varepsilon^m) \right]$$

(die Summen sind wegen V2.3a) alle endlich), folgt die Behauptung mit dem Satz von der majorisierten Konvergenz von Lebesgue.]

S2.8 Satz: Für jedes $b \in \mathbb{R}^+$ gibt es eine Zahl $m_b \in \mathbb{N}$, die nur von b abhängt, so daß für alle $m \geq m_b$ und alle $\bar{q}$ gilt

$$\inf_{\varepsilon \in \mathcal{E}^m} \bar{r}(\bar{q}, \varepsilon^m) \leq \inf_{\varepsilon \in \mathcal{E}} \bar{r}(\bar{q}, \varepsilon) + b.$$

Beweis: Sei m die Zahl der Entscheidungen, nach denen eine Aussage aus A getroffen wird. Wk $\{m \geq m_b \mid \bar{q}, \varepsilon\}$ bezeichne die WK dafür, daß $m \geq m_b$ ausfällt, wenn ε benutzt wird und $\bar{q}$ a priori Verteilung über K ist. Sei $v_o \in \mathbb{R}^+$ eine obere Schranke von $v(k; a)$ und $b \in \mathbb{R}^+$, dann gibt es mit V2.4 b) eine kleinste Zahl $m^*(b) \in \mathbb{N}_0$ derart, daß

$$k(v_m, w_m) \geq \frac{v_o^2}{b},$$

wenn nur

$$\max_{i=1, \dots, n} l(u_1^i \dots u_m^i) \geq m^*(b).$$

Sei $\varepsilon_1 \in \mathcal{E}$ beliebig festgelegt und $m_b := n \cdot m^*(b)$, dann gilt entweder

$$1) \text{ Wk } \{m \geq m_b \mid \bar{q}, \varepsilon_1\} \geq \frac{b}{v_o} \quad \text{oder}$$

$$2) \text{ Wk } \{m \geq m_b \mid \bar{q}, \varepsilon_1\} < \frac{b}{v_o}.$$

Gelte 1). Dann ist offensichtlich $\bar{r}(\bar{q}, \varepsilon_1) \geq v_o$. Dann werde ε_2 so gewählt, daß ohne eine Beobachtung eine Aussage aus A getroffen wird und es ist $\bar{r}(\bar{q}, \varepsilon_2) \leq v_o \leq \bar{r}(\bar{q}, \varepsilon_1)$.

Gelte 2). Wegen K2.4 gibt es in $\mathcal{E}$ ein ε_2 , so daß $\varepsilon_2(\cdot \mid v_m, w_m) = \varepsilon_1(\cdot \mid v_m, w_m)$, wenn nur $\max_{i=1, \dots, n} l(u_1^i \dots u_m^i) < m^*(b)$

und sonst $\varepsilon_2(a_o \mid v_m, w_m) = 1$ für ein bestimmtes $a_o \in A$.

Offensichtlich gilt für ε_2 und die Zahl m der Entscheidungen, nach denen eine Aussage aus A getroffen wird, die Abschätzung $m \leq n \cdot \max_{i=1, \dots, n} l(u_1^i \dots u_m^i) < n \cdot m^*(b) = m_b$.

Da aber 2) gilt, folgt

$$\bar{r}_1(\bar{q}, \varepsilon_2) \leq \bar{r}_1(\bar{q}, \varepsilon_1) + v_0 \frac{b}{v_0} \quad \text{und}$$

$$\bar{r}_2(\bar{q}, \varepsilon_2) \leq \bar{r}_2(\bar{q}, \varepsilon_1).$$

In beiden Fällen gilt $\varepsilon_2 \in \mathcal{E}^{m_b}$ (vergl. p. 25) und

$$\bar{r}(\bar{q}, \varepsilon_2) \leq \bar{r}(\bar{q}, \varepsilon_1) + b.$$

Daraus folgt die Behauptung des Satzes.]

Mit dem Beweis des letzten Satzes ist es gelungen, die formale Analogie zur Theorie der statistischen Entscheidungsfunktionen, die in [18] niedergelegt worden ist, herzustellen. Man erkennt sofort (die in dieser Arbeit angegebenen Begriffe sind gerade in geeigneter Weise definiert worden), welche Definitionen sich gegenseitig entsprechen; die Korrespondenzen brauchen dann nicht mehr expliziert angegeben zu werden. Die den ER interessierenden Tatsachen werden im folgenden zitiert, jedoch die Beweise nicht mehr angegeben. Man könnte diese ohne Zusatzüberlegungen aus [18] übernehmen.

Sei $\mathcal{M}'(K)$ die Menge aller W_K -Maße über $\mathcal{X}_K$, der $\mathcal{G}$ -Algebra über K .

S2.9: Das Entscheidungsproblem ist streng determiniert, d.h.

$$\sup_{\bar{q} \in \mathcal{M}'(K)} \inf_{\varepsilon \in \mathcal{E}} \bar{r}(\bar{q}, \varepsilon) = \inf_{\varepsilon \in \mathcal{E}} \sup_{\bar{q} \in \mathcal{M}'(K)} \bar{r}(\bar{q}, \varepsilon).$$

S2.10: Für alle $\bar{q} \in \mathcal{M}'(K)$ existiert eine Bayes-EV $\varepsilon_{\bar{q}} \in \mathcal{E}$ bzgl. $\bar{q}$ (vergl. (2.15)).

S2.11: Es gibt eine Minimax-EV $\varepsilon' \in \mathcal{E}$ (vergl. (2.16)).

S2.12: Sei $\bar{q}' \in \mathcal{M}^1(K)$ am wenigsten günstig (vergl. (2.17)), dann in jede Minimax-EV eine Bayes-EV bzgl. $\bar{q}'$.

S2.13: Sei $\bar{q}' \in \mathcal{M}^1(K)$ am wenigsten günstig und $\varepsilon' \in \mathcal{E}$ Minimax-EV sowie $K' \subset K$ die Menge aller k' , so daß

$$r(k', \varepsilon') < \sup_{k \in K} r(k, \varepsilon'),$$

dann ist $\bar{q}'(K') = 0$.

Sei auf $K \times K$

$$d_{\mathcal{E}}(k_1, k_2) := \sup_{\varepsilon \in \mathcal{E}} |r(k_1, \varepsilon) - r(k_2, \varepsilon)|.$$

$d_{\mathcal{E}}$ induziert auf K eine offene Umgebungstopologie.

S2.14: Sei $\bar{q}' \in \mathcal{M}^1(K)$ am wenigsten günstig und $\varepsilon' \in \mathcal{E}$ Minimax-EV, dann ist

$$r(k', \varepsilon') = \max_{k \in K} r(k, \varepsilon')$$

für alle k' , für die es keine bzgl. $d_{\mathcal{E}}$ offene Menge $K' \subset K$ mit $k' \in K'$ gibt, so daß $\bar{q}'(K') = 0$.

Man sagt, daß ε^{m+1} aus $\varepsilon \in \mathcal{E}$ durch Abschneiden nach der m-ten Entscheidung erzeugt wird, wenn für alle $s \in \mathbb{N}$ mit $s < m$ gilt $\varepsilon^{m+1}(\cdot | v_s, w_s) = \varepsilon(\cdot | v_s, w_s)$ und

$\varepsilon^{m+1}(A | v_m, w_m) = 1$. Also wird bei Benutzung von ε^{m+1} nach spätestens $m+1$ Entscheidungen eine Aussage aus A getroffen worden sein, d.h. $\varepsilon^{m+1} \in \mathcal{E}^{m+1}$ (vergl. p. 25).

S2.15: Wird ε^{m+1} aus $\varepsilon \in \mathcal{E}$ für alle $m \in \mathbb{N}_0$ durch Abschneiden nach der m-ten Entscheidung gewonnen, so gilt für alle $\bar{q} \in \mathcal{M}^1(K)$

$$\lim_{m \rightarrow \infty} \bar{r}(\bar{q}, \varepsilon^m) = \bar{r}(\bar{q}, \varepsilon).$$

Definiert man auf $\mathcal{M}'(K) \times \mathcal{M}'(K)$ für $m \in \mathbb{N}$

$$d'_m(\bar{q}_1, \bar{q}_2) := \sup_{\varepsilon^m \in \mathcal{E}^m} |\bar{r}(\bar{q}_1, \varepsilon^m) - \bar{r}(\bar{q}_2, \varepsilon^m)|,$$

dann induziert d'_m eine offene Umgebungstopologie in $\mathcal{M}'(K)$.

S2.16: Es gibt eine Folge $(\bar{q}_i)_{i \in \mathbb{N}}$ aus $\mathcal{M}'(K)$ derart, daß für jedes $m \in \mathbb{N}$ die Menge $\{\bar{q}_i : i \in \mathbb{N}\}$ bzgl. d'_m dicht in $\mathcal{M}'(K)$ liegt.

S2.17: $(\bar{q}_i)_{i \in \mathbb{N}}$ sei eine Folge aus $\mathcal{M}'(K)$ derart, daß für jedes $m \in \mathbb{N}$ $\{\bar{q}_i : i \in \mathbb{N}\}$ bzgl. d'_m dicht in der Menge $\{\bar{q}_i : i \in \mathbb{N}\} \cup \{\bar{q}_k : k \in K, \bar{q}_k \in \mathcal{M}'(K), \bar{q}_k(\langle k \rangle) = 1\}$ liegt. Dann gibt es eine Minimax-EV ε' und eine Folge $(\varepsilon_i)_{i \in \mathbb{N}}$ aus $\mathcal{E}$, die regulär gegen ε' konvergiert, und jedes ε_i ist Bayes-EV bzgl. eines $\bar{q}'_i \in \mathcal{M}'(K)$, wobei $\bar{q}'_i$ Linearkombination einer endlichen Zahl von Elementen aus $\{\bar{q}_i : i \in \mathbb{N}\}$ ist.

Wenn $\{k_i : i \in \mathbb{N}\}$ für jedes $m \in \mathbb{N}$ bzgl. d'_m (vergl. (2.33)) dicht in K liegt, dann ist die Folge $(\bar{q}_{k_i})_{i \in \mathbb{N}}$ eine Folge, die die Voraussetzung von S2.17 erfüllt.

Um die weiter unten angegebenen Sätze beweisen zu können, benötigt man zwei zusätzliche Voraussetzungen, die nun formuliert werden und im folgenden gelten sollen:

V2.1 c) Die Folge $(k_i)_{i \in \mathbb{N}_0}$ aus K konvergiere regulär gegen k_0 , dann konvergiere $(v(k_i, a))_{i \in \mathbb{N}_0}$ gleichmäßig in a gegen $v(k_0, a)$.

V2.6 K ist kompakt bzgl. der regulären Topologie.

Schon nach den Bemerkungen über die die reguläre Topologie in K ist K lokal-kompakt und besitzt K eine abzählbare Basis. Demnach ist es möglich, die schwache Topologie in $\mathcal{M}'(K)$ einzuführen.

S2.18: Eine Folge $(\bar{q}_i)_{i \in \mathbb{N}_0}$ aus $\mathcal{M}'(\mathcal{K})$ konvergiere schwach gegen $\bar{q}_0$, dann gilt

$$a) \lim_{i \rightarrow \infty} \bar{r}(\bar{q}_i, \varepsilon^m) = \bar{r}(\bar{q}_0, \varepsilon^m) \quad \text{gleichmäßig in } \varepsilon^m$$

für $m \in \mathbb{N}$,

$$b) \lim_{i \rightarrow \infty} \inf_{\varepsilon} \bar{r}(\bar{q}_i, \varepsilon) = \inf_{\varepsilon} \bar{r}(\bar{q}_0, \varepsilon),$$

$$c) \lim_{i \rightarrow \infty} \bar{r}(\bar{q}_i, \varepsilon) \geq \bar{r}(\bar{q}_0, \varepsilon) \quad \text{für alle } \varepsilon \in \mathcal{E}.$$

S2.19: Es gibt eine am wenigsten günstige a priori Verteilung $\bar{q}' \in \mathcal{M}'(\mathcal{K})$ (vergl. (2.17)).

Also gilt auch, daß jede Minimax-EV Bayes-EV bzgl. jeder am wenigsten günstigen a priori Verteilung ist.

S2.29: Der reguläre Grenzwert einer Folge von Bayes-EV ist selbst Bayes-EV.

S2.21: Sei $\mathcal{E}^b \subset \mathcal{E}$ die Menge aller EV ε mit der Eigenschaft, daß $r(k, \varepsilon)$ als Funktion von k beschränkt bleibt. Dann ist die Menge aller Bayes-EV vollständig relativ zu $\mathcal{E}^b$ (vergl. (2.19)).

2.2 Spezialisierungen durch Wahl der Kostenfunktion

Für die folgenden Überlegungen sei einmal angenommen, die Zahlenfolge $(b_s)_{s \in \mathbb{N}_0}$ aus V2.3 a) sei festgelegt. Dann hängt die Menge $\mathcal{E}$ aller EV, die dem ER zur Verfügung steht, wegen der Bedingungen V2.3 b) und V2.4 b) allein von der Wahl der Kostenfunktion k ab. Die Abhängigkeit wird im vorliegenden Abschnitt durch verschiedene Wahlen für die Kostenfunktion beleuchtet. Dieses Vorgehen führt zu einigen Definitionen von speziellen Formen von EV und dazugehörigen Experimenten. Insbesondere läßt sich

bei nichtadaptiven Experimenten eine interessante notwendige und hinreichende Bedingung für eine Bayes-EV angeben. Durch eine weitere Spezialisierung der nicht adaptiven Experimente gelangt man zu den Vorgabeexperimenten, die im Abschnitt 2.3 ausführlicher behandelt werden.

2.2.1 Buchstabenweise adaptive Experimente

Zunächst sei angenommen, daß die Kostenfunktion nur von den durch Konkatenation entstandenen Wörtern $u_1^i \dots u_m^i$ und $v_1^i \dots v_m^i$ ($i = 1, \dots, n$) und nicht zusätzlich noch von den in v_m bzw. w_m definierten Gruppierungen von Wörtern ($u_1^i, \dots, u_m^i$) bzw. ($v_1^i, \dots, v_m^i$) ($i = 1, \dots, n$) abhängt. Man kann also schreiben

$$(2.37) \quad k(v_m, w_m) = k(u_1^i \dots u_m^i, v_1^i \dots v_m^i \quad (i=1, \dots, n)).$$

Eine EV ε^* operiert buchstabenweise adaptiv, wenn sie folgender Bedingung genügt:

Für alle $m \in \mathbb{N}_0$, alle v_m und w_m mit $\ell(w_m) = \ell(v_m)$ gilt

$$\sum_{x^1 \in X^*} \dots \sum_{x^n \in X^*} \varepsilon^*(x^1, \dots, x^n \mid v_m, w_m) + \varepsilon^*(A \mid v_m, w_m) = 1.$$

Ein dazugehöriges Experiment heißt ebenfalls buchstabenweise adaptiv und hat fast sicher die Eigenschaft, daß bei einer Entscheidung entweder ein $a \in A$ oder ein $(x^1, \dots, x^n) \in \bigcap_{i=1}^n X^*$ anstatt wie im allgemeinen Falle ein Element aus $\bigcap_{i=1}^n X^*$ gewürfelt wird.

Über den Zusammenhang von buchstabenweise adaptiven EV und den oben erklärten Kostenfunktionen gibt der folgende Satz Auskunft.

S2.22 Satz: Gelte (2.37), dann gibt es zu jedem $\varepsilon \in \mathcal{E}$ eine buchstabenweise adaptive EV ε^* mit der Eigenschaft: Für alle $k \in K$ ist

$$r(k, \varepsilon) \geq r(k, \varepsilon^*).$$

Beweis: Sei für $A' \in \mathcal{X}_A$, $u^i \in X^*$, $v^i \in Y^*$ mit $l(v^i) = l(u^i)$ ($i = 1, \dots, n$)

$$(2.38) \quad p_z(u_1^1, \dots, u_n^1, v_1^1, \dots, v_n^1, A' | k, \varepsilon) := \sum_{m, \omega_m: u_1^1 \dots u_n^1 = u^i \ (i=1, \dots, n)} p_z(\omega_m, \omega_m, A' | k, \varepsilon)$$

(vergl. (2.8)). Summiert wird also über alle möglichen Gruppierungen in ω_m , so daß die Konkatenation der Worte $u_1^1, \dots, u_n^1$ gerade u^i ergibt. Diese Schreibweise ist sinnvoll, da in p_z stets $l(\omega_m) = l(\omega_m)$ vorausgesetzt wird. Dann ist mit (2.37) und (2.12)

$$(2.39) \quad r_2(k, \varepsilon) = \sum_{u_1^1, \dots, u_n^1} \sum_{v_1^1, \dots, v_n^1, l(v^i) = l(u^i)} k(u_1^1, v_1^1, \dots, u_n^1, v_n^1) \cdot p_z(u_1^1, \dots, u_n^1, v_1^1, \dots, v_n^1, A' | k, \varepsilon).$$

Eine gewünschte buchstabenweise adaptive EV ε^* wird aus ε folgendermaßen konstruiert: Sei (in der ursprünglichen Schreibweise) für $A' \in \mathcal{X}_A$

$$1) \quad \varepsilon^*(A' | u_1^1, \dots, u_n^1, v_1^1, \dots, v_n^1) := \varepsilon(A' | u_1^1, \dots, u_n^1, v_1^1, \dots, v_n^1),$$

2) Sei

$$u^i := x_1^i \dots x_m^i, \quad v^i := y_1^i \dots y_m^i \quad (i=1, \dots, n), \quad \mathcal{U}_m := (x_j^i), \quad \mathcal{V}_m := (y_j^i)$$

und $x_j^i \in X$ sowie $y_j^i \in Y$. Dann wird mit (2.8) definiert

$$(2.40) \quad p_2(\mathcal{U}_m, \mathcal{V}_m, A' | k, \varepsilon^*) := p_z(u_1^1, \dots, u_n^1, v_1^1, \dots, v_n^1, A' | k, \varepsilon).$$

p_z läßt sich aber darstellen

$$p_z(\mathcal{U}_m, \mathcal{V}_m, A' | k, \varepsilon^*) = \prod_{j=0}^{m-1} \varepsilon^*(x_{j+1}^1, \dots, x_{j+1}^n | \mathcal{U}_j, \mathcal{V}_j) \cdot \varepsilon^*(A' | \mathcal{U}_m, \mathcal{V}_m) k(\mathcal{V}_m | \mathcal{U}_m).$$

k sei als bekannt vorausgesetzt und $\varepsilon^*(A' | \mathcal{U}_m, \mathcal{W}_m)$ ist in 1) definiert worden. Damit läßt sich ε^* durch Divisionen aus der letzten Gleichung bis auf Argumente, bei denen der Dividend Null wird (diese spielen jedoch bei der Bestimmung des Risikos keine Rolle), berechnen.

Wegen (2.39) ist $r_2(k, \varepsilon) = r_2(k, \varepsilon^*)$. Aus (2.38), (2.40), (2.9), (2.10) und (2.11) folgert man, daß auch $r_1(k, \varepsilon) = r_1(k, \varepsilon^*)$ gilt.

Hat k also die Form (2.37), so braucht man nur noch buchstabenweise adaptive EV zu betrachten. Die Menge $\mathcal{E}$ kann dann auf die Menge $\mathcal{E}_{ba}$ der buchstabenweise adaptiven Experimente eingeschränkt werden. Die Existenz von Bayes-, Minimax-EV, am wenigsten günstigen a priori Verteilungen sowie deren Zusammenhänge, die in 2.1.3 für $\mathcal{E}$ erläutert wurden, bleiben dann auch in $\mathcal{E}_{ba}$ erhalten.

2.2.2 Nichtadaptive Experimente

Die Kostenfunktion k möge in diesem Abschnitt folgende Eigenschaft haben:

Für alle $m \geq 2$ und v_m sei $k(v_m, \cdot) = \infty$.

Aus dieser Bedingung folgt, daß nach spätestens einer Eingabe in die Kopien des Kanales eine Aussage aus A getroffen wird. Also stimmt $\mathcal{E}$ mit der Menge $\mathcal{E}^2$ überein. Die Elemente von $\mathcal{E}^2$ heißen nichtadaptive-EV und die zugehörigen Experimente nichtadaptiv. Jede EV ε aus $\mathcal{E}^2$ hat die Eigenschaft:

Wenn $v_1 \neq (\Lambda_1, \dots, \Lambda)$ ist, folgt

$$\varepsilon(A | v_1, \cdot) = 1.$$

Mit dem Satz von der monotonen Konvergenz von Levi ist das bedingte Risiko (vergl. (2.13))

$$(2.41) \quad r(k, \varepsilon) = \int_A r(k, a) \varepsilon(da | v_0, w_0) \\ + \sum_{v_1} \varepsilon(v_1 | v_0, w_0) \sum_{w_1} k(w_1 | v_1) \left[\int_A r(k, a) \varepsilon(da | v_1, w_1) \right. \\ \left. + k(v_1, w_1) \right].$$

Für das mittlere Risiko (2.14) findet man mit dem Satz von Levi und dem Satz von Fubini

$$(2.42) \quad \bar{r}(\bar{q}, \varepsilon) = \int_A \int_K v(k, a) \bar{q}(dk) \varepsilon(da | v_0, w_0) \\ + \sum_{v_1} \varepsilon(v_1 | v_0, w_0) \sum_{w_1} \left[\int_A \int_K v(k, a) k(w_1 | v_1) \bar{q}(dk) \varepsilon(da | v_1, w_1) \right. \\ \left. + h(v_1, w_1) \int_K k(w_1 | v_1) \bar{q}(dk) \right].$$

In dem folgenden Absatz wird versucht, $\inf_{\varepsilon} \bar{r}(\bar{q}, \varepsilon)$ zu bestimmen. Dazu sind einige Vorbemerkungen notwendig. Sei $(a_i)_{i \in \mathbb{N}}$ eine Folge aus A mit der Eigenschaft

$$\inf_{a \in A} \int_K v(k, a) k(w_1 | v_1) \bar{q}(dk) = \lim_{i \rightarrow \infty} \int_K v(k, a_i) k(w_1 | v_1) \bar{q}(dk) \\ = \int_K \lim_{i \rightarrow \infty} v(k, a_i) k(w_1 | v_1) \bar{q}(dk)$$

(die letzte Gleichheit mit dem Satz von Lebesgue). Da A kompakt bzgl. d_A ist, enthält (a_i) eine gegen ein $a_0 \in A$ konvergente Teilfolge. Wegen der Stetigkeit von v bzgl. d_A gilt dann

$$\inf_{a \in A} \int_K v(k, a) k(w_1 | v_1) \bar{q}(dk) = \int_K v(k, a_0) k(w_1 | v_1) \bar{q}(dk)$$

und das Minimum existiert. Folglich gilt

$$(2.43) \quad \inf_{\varepsilon} \bar{r}(\bar{q}, \varepsilon) = \min \left\{ \min_{a \in A} \int_K v(k, a) \bar{q}(dk), \right. \\ \left. \min_{v_1} \sum_{w_1} \left[\min_{a \in A} \int_K v(k, a) k(w_1 | v_1) \bar{q}(dk) \right. \right. \\ \left. \left. + h(v_1, w_1) \int_K k(w_1 | v_1) \bar{q}(dk) \right] \right\},$$

da $\min_{v_1} \sum_{w_1} \dots$ existiert, weil nach V2.4 b)

$$\sum_{w_1} h(v_1, w_1) \int_K k(w_1 | v_1) \bar{q}(dk) \geq b'_s,$$

wenn nur $\max_{i=1, \dots, n} l(u_i) \geq s$ und weil $\lim_{s \rightarrow \infty} b'_s = \infty$.

Nach der obigen Überlegung ist $\int_K v(k, a) k(\omega, |v_1) \bar{q}(dk)$ eine bzgl. $\mathcal{A}_A$ stetige Funktion von a . Wenn also für ein $a \in A$

$$(2.44) \min_{a' \in A} \int_K v(k, a') k(\omega, |v_1) \bar{q}(dk) < \int_K v(k, a) k(\omega, |v_1) \bar{q}(dk),$$

dann gibt es eine maximale bzgl. $\mathcal{A}_A$ offene Umgebung $U(a)$ von a , so daß die Elemente von $U(a)$ (2.44) erfüllen.

Damit erhält man den folgenden Satz:

S2.23 Satz: Eine EV $\varepsilon \in \mathcal{E}^2$ ist genau dann Bayes-EV bzgl. der a priori Verteilung $\bar{q}$ aus $\mathcal{M}'(K)$, wenn die drei folgenden Bedingungen erfüllt sind: (vergl. (2.43))

$$1) \text{ Falls } \min_{v_1} \sum_{\omega_1} [\dots] < \min_a \int_K v(k, a) \bar{q}(dk),$$

$$\text{dann } \varepsilon(A | v_0, \omega_0) = 0.$$

$$2) \text{ Falls für ein } v_1' \quad \min_{v_1} \sum_{\omega_1} [\dots]$$

$$< \sum_{\omega_1} \left[\min_a \int_K v(k, a) k(\omega_1 | v_1') \bar{q}(dk) + k(v_1', \omega_1) \int_K k(\omega_1 | v_1') \bar{q}(dk) \right],$$

$$\text{dann } \varepsilon(v_1' | v_0, \omega_0) = 0.$$

$$3) \text{ Falls für ein } a \in A \text{ (2.44) gilt,}$$

$$\text{dann } \varepsilon(U(a) | v_1, \omega_1) = 0.$$

Die Bedingung 3) aus S2.23 läßt sich noch etwas umformen.

Sei

$$\bar{v}(\bar{q}, a) := \int_K v(k, a) \bar{q}(dk)$$

und $\bar{q}_{\omega_1 | v_1}$ die a posteriori Verteilung in K , nachdem v_1 eingegeben und ω_1 beobachtet wurde, d.h. für $k' \in \mathcal{K}_K$ ist

$$\bar{q}_{\omega_1 | v_1}(k') = \frac{\int_{K'} k(\omega_1 | v_1) \bar{q}(dk)}{\int_K k(\omega_1 | v_1) \bar{q}(dk)}.$$

Wenn $A(\bar{q}, \omega_1, v_1) \subset A$ die Menge aller α ist, für die

$$\bar{v}(\bar{q}, \omega_1, v_1, \alpha) = \min_{\alpha'} \bar{v}(\bar{q}, \omega_1, v_1, \alpha')$$

(diese Menge ist offensichtlich aus $\mathcal{L}_A$), dann heißt eine zu 3) äquivalente Bedingung

$$3') \quad \int_{A(\bar{q}, \omega_1, v_1)} \varepsilon(d\alpha | v_1, \omega_1) = 1.$$

An dieser Stelle sei bemerkt, daß zur Ableitung von (2.43) und S2.23 die Voraussetzung V2.3 a) nicht benutzt worden ist. Jedoch läßt sich ohne diese oder eine ähnliche Bedingung nicht der grundlegende Satz S2.5 beweisen. Man betrachte dazu nur die Folge von EV $(\varepsilon_i)_{i \in \mathbb{N}}$ mit

$$\varepsilon_i(A | v_0, \omega_0) = 0 \quad (i \in \mathbb{N})$$

und eine Folge $(v_1^i)_{i \in \mathbb{N}}$ verschiedener Elemente aus $\bigcup_{i=1}^n X^*$ mit der Eigenschaft

$$\varepsilon_i(v_1^i | v_0, \omega_0) = 1 \quad (i \in \mathbb{N}).$$

Dann gilt nämlich

$$\lim_{i \rightarrow \infty} \varepsilon_i(A | v_0, \omega_0) = 0,$$

$$\lim_{i \rightarrow \infty} \varepsilon_i(v_1 | v_0, \omega_0) = 0 \quad \text{für jedes } v_1.$$

Dies trifft auch für jede Teilfolge von (ε_i) zu, eine solche also nicht regulär in $\mathcal{E}$ konvergieren kann.

Auf V2.4 b) oder eine analoge Bedingung mit der Abschwächung $\lim_{s \rightarrow \infty} b_s' > \sup_{k, \alpha} v(k, \alpha)$ kann im Falle des Wegfalles von V2.3 a) nicht verzichtet werden. Der Wert (vergl. (2.43))

$$\inf_{v_1} \sum_{\omega_1} \left[\min_{\alpha} \int_{\mathcal{K}} v(k, \alpha) k(\omega_1 | v_1) \bar{q}(dk) + h(v_1, \omega_1) \int_{\mathcal{K}} k(\omega_1 | v_1) \bar{q}(dk) \right]$$

würde dann i.a. nicht für ein geeignetes v_1 angenommen werden, so daß dann die Existenz von Bayes-EV nicht mehr gesichert wäre.

Damit alle Sätze aus 2.1.3 auch für nichtadaptive EV zutreffen, muß man alle wesentlichen Forderungen aus 2.1.2 und 2.1.3 beibehalten. In diesem Fall kann man die Aussage S2.21 noch etwas verschärfen. Wegen V2.3 a) gibt es nur endlich viele v_i , die mit positiver W_k als Eingaben gewählt werden können. Damit ist für jedes $\varepsilon \in \mathcal{E}$ $r(k, \varepsilon)$ eine beschränkte Funktion von k . Dann gilt

S2.24 Satz: Die Menge aller Bayes-EV in der Menge $\mathcal{E}^2$ aller nichtadaptiven EV ist vollständig relativ zu $\mathcal{E}^2$.

Für ein festes $\bar{q} \in \mathcal{M}'(K)$ betrachte man die Menge aller Bayes EV bzgl. $\bar{q}$. Diese EV ε' mögen alle die Bedingung $\varepsilon'(A | v_0, w_0) = 0$

erfüllen. Dann gibt es nach S2.23 stets solche EV ε'' in dieser Menge mit der folgenden Eigenschaft:

Es gibt genau ein v_i' , so daß

$$(2.45) \quad \varepsilon''(v_i' | v_0, w_0) = 1.$$

Eine EV mit der Eigenschaft (2.45) nennt man, falls der ER das v_i' kennt, eine Vorgabe-EV und ein zugehöriges ET ein Vorgabe-ET. Eine solche EV ist also vollständig durch $\varepsilon(\cdot | v_i', \cdot)$ definiert.

Man kann auch durch Spezialisierung der Kostenfunktion zu den Vorgabe-EV gelangen, wenn nämlich gilt:

Es gibt genau ein v_i' , so daß

$$k(v_m, \cdot) = \infty \quad (v_m \neq v_i'),$$

$$k(v_i', \cdot) = 0.$$

Dann ist $\mathcal{E}$ eine spezielle Menge von nichtadaptiven EV. Es gilt aber, daß es zu jedem $\varepsilon \in \mathcal{E}$ eine Vorgabe-EV ε' gibt, so daß für alle $k \in K$

$$r(k, \varepsilon) \geq r(k, \varepsilon'),$$

da nach Voraussetzung eine Beobachtung nichts kostet und der minimale mittlere erwartete Verlust nach einer Beobachtung nicht größer als bei einer a priori Entscheidung ist. Von Interesse ist also nur noch die Menge aller Vorgabe-EV. Den Vorgabe-EV wird in den folgenden Paragraphen die gesamte Aufmerksamkeit gewidmet werden.

2.3 . Vorgabexperimente

In diesem Abschnitt werden die wesentlichen und die im weiteren interessierenden allgemeinen Tatsachen über die am Ende des letzten Abschnittes eingeführten Vorgabe-EV zusammengefaßt und der wichtige Spezialfall, daß die Verlustfunktion eine Fehlerfunktion ist, untersucht. Es sei jetzt angenommen, die Eingabe v_1 ist fest gewählt.

Da die Eingabe von v_1 und die Beobachtung der Reaktion der Kopien des Kanales nichts kostet, ist nach (2.41) das bedingte Risiko

$$(2.46) \quad r(k, \epsilon) = \sum_{w_1} k(w_1 | v_1) \int_A v(k, a) \epsilon(da | v_1, w_1)$$

und nach (2.42) das mittlere Risiko

$$(2.47) \quad \bar{r}(\bar{q}, \epsilon) = \sum_{w_1} \int_A \int_K v(k, a) k(w_1 | v_1) \bar{q}(dk) \epsilon(da | v_1, w_1).$$

Zunächst sind einige Bemerkungen über die Verwirklichung der Forderungen aus 2.1.2 und 2.1.3 anzubringen. Von Interesse für Vorgabe-EV sind bis auf die Meßbarkeitsforderungen V5 nur noch die Voraussetzungen V2.1, V2.2 und V2.6, die die Stetigkeitsfragen von v sowie die Kompaktheitseigenschaften von A und K betreffen. Drei Spezialfälle sind hinsichtlich Experimenten mit Kanälen interessant:

- 1) A und K sind endliche Mengen. Dann benötigt man nur noch die triviale Forderung, daß man mit Hilfe von v die Aussagen aus A unterscheiden kann, d.h. für zwei verschiedene $a_1, a_2 \in A$ existiert ein $k \in K$ mit der Eigenschaft $v(k, a_1) \neq v(k, a_2)$. Alle weiteren Annahmen sind dann von selbst erfüllt.

Z.B. haben alle zustandsbestimmenden Experimente mit endlichen Automaten diese Form. Die Menge Z der Zustände ist dann eine Indexmenge und K läßt sich darstellen als $\{k_z : z \in Z\}$. A ist in diesem speziellen Fall eine endliche Menge von Aussagen, die die Zustände des Automaten betreffen. Am Ende dieses Abschnittes und in Kapitel 3 wird weiter auf solche Experimente eingegangen.

- 2) K wird beschrieben durch eine endlich parametrische Menge G , die insbesondere eine Teilmenge des $\mathbb{R}^s$ mit $s \in \mathbb{N}$ sein soll. Dann hat K die Form $\{k_g : g \in G\}$. A ist endlich. Zunächst muß wieder V2.2 a), d.h. die Unterscheidbarkeit der Aussagen, gefordert werden. Setzt man weiter voraus:

- a) k_g ist eine (im ungewöhnlichen Sinne) stetige Funktion von g ,
- b) $v(k_g, a)$ ist für jedes a eine stetige Funktion von g ,
- c) G ist abgeschlossen und beschränkt in $\mathbb{R}^s$,

dann sind offensichtlich alle Forderungen erfüllt.

Setzt man nur V2.2 a) und die Beschränktheit von $v(k_g, a)$ bzgl. g voraus, dann gelten schon die Forderungen aus 2.1.2 bis auf V2.1 c) und V2.6.

Dieser Fall tritt z.B. auf, wenn Anfangsverteilungen der Zustände eines endlichen Automaten vorliegen, man also K wie folgt darstellen kann: wenn $k \in K$, dann gibt es ein $q^+ \in \mathcal{M}^+(Z)$ und es gilt $k = \sum_{z \in Z} q^+(z) k_z$ wobei k_z wie in 1) festgelegt ist. A kann dann aus

endlich vielen Aussagen über die Menge der zulässigen q^+ bestehen.

- 3) K ist wie in 2) durch eine endlich parametrische Menge $G \subset \mathbb{R}^s$ festgelegt und A ist gleich G (was damit gemeint wird, ist klar). Mit V2.2 a) fordert man wieder a) und c) aus 2) sowie

b') $v(k_g, g') \quad (g, g' \in G)$ ist eine stetige Funktion von g und g' .

Damit sind alle benötigten Voraussetzungen erfüllt.

Der hier behandelte Sachverhalt trifft z.B. bei Punktabschätzungen zu. Wenn wie in 2) $k = \sum_{z \in \mathbb{Z}} q^+(z) k_z$

ist, dann kann das q^+ geschätzt werden.

Bei Intervallabschätzungen geht man so ähnlich vor.

Man wählt K wie oben und $A = G \times H$, wobei $H \subset \mathbb{R}^{s'}$ mit $s' \in \mathbb{N}$ und ein zulässiges Intervall durch seinen Mittelpunkt g und durch ein $h \in H$ vollständig beschrieben wird. Man fordert wieder V2.2 a), a) und

b'') $v(k_g, (g', h))$ ist eine stetige Funktion von g , g' und h ,

c') $G \times H$ ist abgeschlossen und beschränkt in $\mathbb{R}^{s+s'}$.

Wenn für ein $k \in K$ wieder gilt $k = \sum_{z \in \mathbb{Z}} q^+(z) k_z$,

dann schätzt man, ob q^+ in dem Intervall liegt, das durch seinen Mittelpunkt g und durch die Parameter h beschrieben wird.

Bei Annahme der Voraussetzungen aus 2.1.2 und 2.1.3 sind die wichtigsten bis zu dieser Stelle bekannten Tatsachen über Vorgabe EV:

Für jede a priori Verteilung $\tilde{q}$ gibt es eine Bayes-EV bzgl. $\tilde{q}$. Notwendig und hinreichend dafür, daß eine EV eine Bayes-EV bzgl. $\tilde{q}$ darstellt, ist die Gültigkeit der Bedingung 3) (oder 3')) aus S2.23.

Es gibt eine Minimax-EV.

Es gibt eine am wenigsten günstige a priori Verteilung. Jede Minimax-EV ist Bayes-EV bzgl. jeder am wenigsten günstigen a priori Verteilung.

Wenn $\bar{q}$ am wenigsten günstig und ε Minimax-EV ist, dann gilt $\bar{q}\{k: r(k, \varepsilon) = \max_{k' \in K} r(k', \varepsilon)\} = 1$.

Die Menge aller Bayes-Vorgabe-EV ist vollständig relativ zur Menge aller Vorgabe-EV.

Zum Schluß dieses Abschnittes wird ein schon angekündigter Spezialfall näher beleuchtet. Sei $K := \{k_i : i \in \{1, \dots, t\}\}$ mit $t \in \mathbb{N}$ und $A := \{1, \dots, t\}$. Weiterhin möge die Verlustfunktion eine Fehlerfunktion der Form $v(k_i, j) = 1 - \delta_{ij}$ sein. Dann vereinfachen sich (2.46) und (2.47) zu

$$(2.48) \quad r(k_i, \varepsilon) = 1 - \sum_{\omega_1} k_i(\omega_1 | v_1) \varepsilon(i | v_1, \omega_1)$$

(bedingter Entscheidungsfehler),

$$(2.49) \quad \bar{r}(\bar{q}, \varepsilon) = 1 - \sum_{\omega_1} \sum_{i=1}^t k_i(\omega_1 | v_1) \bar{q}(i) \varepsilon(i | v_1, \omega_1)$$

(mittlerer Entscheidungsfehler mit der Abkürzung $\bar{q}(i) := \bar{q}(k_i)$).

Damit ist notwendig und hinreichend dafür, daß ε Bayes-EV bzgl. $\bar{q}$ ist,

wenn $k_i(\omega_1 | v_1) \bar{q}(i) < \max_{j=1, \dots, t} k_j(\omega_1 | v_1) \bar{q}(j)$,

dann $\varepsilon(i | v_1, \omega_1) = 0$.

Eine Minimax-EV kann man finden, wenn man eine am wenigsten günstige a priori Verteilung $\bar{q}$ kennt und unter den Bayes-EV ε bzgl. $\bar{q}$ eine solche aussucht, für die

$$r(k_i, \varepsilon) = \max_{j=1, \dots, t} r(k_j, \varepsilon),$$

wenn nur $\bar{q}(i) > 0$ ist.

Man kann das Finden einer am wenigsten günstigen Anfangsverteilung folgendermaßen erleichtern:

Für jedes $\bar{q}$ wähle man folgende Bayes-EV $\varepsilon_{\bar{q}}$ als Repräsentanten:

$\varepsilon_{\bar{q}}(i | v_1, w_1) = 1$, wenn i der kleinste Index ist, für den $k_i(w_1 | v_1)q(i) = \max_j k_j(w_1 | v_1)q(j)$.

Da dann $\bar{v}(\bar{q}, \varepsilon_{\bar{q}})$ nur noch von $\bar{q}$ abhängt, maximiert eine am wenigsten günstige a priori Verteilung gerade $\bar{v}(\bar{q}, \varepsilon_{\bar{q}})$

Hilfreich beim Suchen einer am wenigsten günstigen a priori Verteilung ist der folgende Satz.

S2.24 Satz: Alle Komponenten einer am wenigsten günstigen a priori Verteilung sind positiv, wenn gilt:

- 1) Es gibt eine EV ε mit $r(k_i, \varepsilon) < 1$ für alle i
- 2) Alle w_1 und i erfüllen:
 $k_i(w_1 | v_1) = 0$ impliziert $k_j(w_1 | v_1) = 0$
für alle j .

Beweis: Sei ε' Minimax-EV. Dann gilt wegen 1)

$$(2.50) \quad r(k_i, \varepsilon) < 1 \quad \text{für alle } i.$$

Weiter sei $\bar{q}'$ am wenigsten günstig, dann ist ε' Bayes-EV bzgl. $\bar{q}'$. Angenommen sei $q'(1) = 0$. Dann ist aber

$\varepsilon'(1 | v_1, w_1) = 0$, außer höchstens für solche w_1 , für die $\sum_{i=1}^t \bar{q}'(i) k_i(w_1 | v_1) = 0$. Folglich muß aber wegen 2)

für diese w_1 $k_i(w_1 | v_1) = 0$ für alle i gelten. Dann ist aber notwendig $r(k_i, \varepsilon') = 1$. Das ist ein Widerspruch zu (2.50).

2.4 Vorgabeexperimente bei unvollständiger Kenntnis des Experimentators über das Experiment

Man nehme einmal an, dem ER würde bei einem ET nicht die volle Information über die Eingaben und die Ausgaben der Kanäle zur Verfügung stehen. Dieser Fall wird i.a. nicht mehr mit den bisher vorgestellten Methoden zu behandeln sein. Am Beispiel von in solcher Weise verallgemeinerten Vorgabe-ET wird im folgenden gezeigt, wie man eine dem Abschnitt 2.3 entsprechende Theorie durch geeignete Definitionen aufbauen kann. Einschränkend soll dazu vorausgesetzt werden, daß der ER in den Ausgabewörtern einige Buchstaben nicht lesen kann und daß seine Kenntnis über die Eingabewörter darin besteht, daß diese aus einer festen endlichen Menge stammen und jeweils mit bestimmten Wahrscheinlichkeiten von einem Eingabemechanismus gewählt werden.

Genauer werden die folgenden Vereinbarungen getroffen: Der ER kennt die "Wischfunktion", nach der die Buchstaben in den Ausgabewörtern unleserlich gemacht werden:

$$w: \prod_{i=1}^n Y^{l_i} \rightarrow \prod_{i=1}^n [Y \cup \{\perp\}]^{l_i},$$

wobei $\perp$ Leerstelle bedeutet und $l_i \in \mathbb{N}$ ($i = 1, \dots, n$) feste Wortlängen sind. w setzt für gewisse Buchstaben in den Eingabewörtern das Symbol $\perp$ und läßt den Rest der Wörter unverändert. Von einem Ausgabewort w_1 kennt der ER also nur $w(w_1)$.

Der ER weiß von den Eingabewörtern nur, daß sie Realisierungen einer ZV ξ über $\prod_{i=1}^n X^{l_i}$ ist, die eine i.a. von dem vorliegenden Kanal, der dem Experimentator ja unbekannt ist, abhängende Verteilung besitzt. Ist κ eine ZV über $\mathcal{K}$, dann kennt der ER die $\text{Wk} \{ \xi = v_1, \kappa = k \} =: \tilde{p}(v_1, k)$ für alle $k \in \mathcal{K}$. $\tilde{p}(\cdot | k)$ ist eine Funktion über $\mathcal{K}$

in einem $\mathbb{R}^s$ für ein Gewisses $s \in \mathbb{N}$. Es sei für das Weiteres vorausgesetzt, daß $\tilde{p}(\cdot|k)$ eine bzgl. der regulären Topologie in K stetige Funktion von k sei.

Die EV $\bar{e}$ definiert man jetzt etwas anders als gewohnt: Anstatt von v_1 und w_1 hängen sie von $\tilde{p}$ und $w(w_1)$ ab; $\bar{e}(\cdot|\tilde{p}, w(w_1))$ ist ein Element von $\mathcal{M}^1(A)$ und legt wie bei Vorgabe-EV eine EV im wesentlichen fest.

Der Vollständigkeit halber wird die Kostenfunktion noch festgelegt:

$$k(v_1, \cdot) = \begin{cases} 0 & , \text{ wenn } \ell(v_1) = \ell_1 := (l^1, \dots, l^n) \\ \infty & \text{sonst} \end{cases}$$

Als bedingtes Risiko definiert man dann

$$\begin{aligned} (2.51) \quad r(k, \bar{e}(\cdot|\tilde{p}, w(\cdot))) &= \overline{\sum_{v_1: \ell(v_1) = \ell_1}} \tilde{p}(v_1|k) \overline{\sum_{w_1: \ell(w_1) = \ell_1}} k(w_1|v_1) \\ &\quad \cdot \int_A v(k, a) \bar{e}(da|\tilde{p}, w(w_1)) \\ &= \overline{\sum_{w(w_1)}} \bar{p}(w(w_1)|\tilde{p}, k) \int_A v(k, a) \bar{e}(da|\tilde{p}, w(w_1)), \end{aligned}$$

wobei sich die Summe der letzten Zeile über den gesamten Bildbereich von w erstreckt (ℓ_1 ist festgelegt) und

$$\bar{p}(w(w_1)|\tilde{p}, k) :=$$

$$\overline{\sum_{w'_1: w(w'_1) = w(w_1)}} \overline{\sum_{v_1: \ell(v_1) = \ell_1}} \tilde{p}(v_1|k) k(w'_1|v_1).$$

Dann ergibt sich das mittlere Risiko zu

$$\begin{aligned} (2.52) \quad \bar{r}(\bar{q}, \bar{e}(\cdot|\tilde{p}, w(\cdot))) &= \int_K \bar{q}(dk) r(k, \bar{e}(\cdot|\tilde{p}, w(\cdot))) \\ &= \overline{\sum_{w(w_1)}} \int_A \int_K v(k, a) \bar{p}(w(w_1)|\tilde{p}, k) \bar{q}(dk) \bar{e}(da|\tilde{p}, w(w_1)). \end{aligned}$$

Der Erwartungswert existiert, da mit $\tilde{p}$ auch $\bar{p}(\cdot|\tilde{p}, k)$ eine stetige Funktion von k ist.

Vergleicht man das bedingte Risiko (2.46) für Vorgabeexperimente mit (2.51), so stellt man eine formale Ähnlichkeit der Ausdrücke fest. $\bar{p}$ nimmt in (2.51) die Rolle des Faktors k in (2.46) ein, $\tilde{p}$ die von v_i und der Bildbereich von w die Rolle aller w_i mit $\ell(w_i) = \ell(v_i)$. Die Annahme, daß $\tilde{p}$ stetig von k abhängt, erlaubt nun, analoge Schlußweisen wie in 2.1.3, 2.2.2 und 2.3 durchzuführen, vorausgesetzt im wesentlichen V2.1, V2.2, V2.5 und V2.6 sind erfüllt. Man erhält also eine zu der Theorie der Vorgabe-EV ganz analoge Theorie mit den entsprechenden Ergebnissen.

2.5 Bemerkungen

- 1) Zunächst wird auf einige Probleme aufmerksam gemacht, die bei der Wahl von günstigen EV in praktischen Fällen auftreten können. Angenommen, man betrachtet adaptive EV und der ER kennt das Entscheidungsproblem und zusätzlich die vorliegende a priori Verteilung der Kanäle. Der ER weiß, daß es Bayes-Lösungen gibt, jedoch kann er sie i.a. nicht bestimmen, weil ihm keine Kriterien oder Algorithmen zur Berechnung von Bayes-EV zur Verfügung stehen. Der ER wird häufig a priori Verteilung der Kanäle nicht kennen oder es brauchen gar keine solchen zu existieren. In diesem Fall kann man Minimax-EV als günstige EV ansehen. Der ER weiß wieder von der Existenz von Minimax-EV, kann sie aber nicht berechnen. Nicht einmal bei Vorgabeexperimenten ist das allgemein möglich, obwohl man die Bayes-EV alle angeben kann. Bei der Bestimmung von am wenigsten günstigen a priori Verteilungen, die auf dem Wege zur Berechnung von Minimax-EV hilfreich sind, treten i.a. dieselben Hindernisse auf. Deshalb muß man sich meist mit suboptimalen Lösungen zufrieden geben. Z.B. setzt man die Menge aller zur Verfügung stehenden EV als

eine endlich parametrische Menge an. Dann kann man probieren, in dieser Parametermenge zu optimieren, entweder systematisch oder mit Monte-Carlo-Methoden.

- 2) Zur Erläuterung der oben angedeuteten Schwierigkeiten wird ein Beispiel angeführt: Vorgabeexperimente werden betrachtet, und es soll eine Punktabschätzung vorgenommen werden. Es seien $A = G = [0, 1]$ und $K := \{k_g : k_g = k_1 g + k_2 (1-g), g \in G\}$, wobei k_1 und k_2 dem ER bekannte Kanäle sind. Das Experiment besteht darin, daß dem ER n Kopien eines k_g vorliegen, er $v_i = (v_i^1, \dots, v_i^n)$ in die Kopien eingibt (d.h. in alle Kopien wird dasselbe u eingegeben) und dann aus der Ausgabe $w_i = (v_i^1, \dots, v_i^n)$ mit $l(v_i) = l(u)$ das ihm unbekannte g schätzt. Die Verlustfunktion sei $v(k_g, g') := (g - g')^2$. Offensichtlich braucht man nur noch Entscheidungsfunktionen der Form $\varepsilon(g' | v_i, w_i) \in \{0, 1\}$ zu betrachten. Dann kann man für das bedingte Risiko schreiben

$$r(k_g, g') = \mathbb{E}_{k_g} (g - g')^2 = \sum_{w_i} (g - g'(w_i))^2 k_g(w_i | v_i)$$

und für das mittlere Risiko, wenn der ER $\bar{q}$ kennt,

$$\bar{r}(\bar{q}, g') = \sum_{w_i} \int_{[0,1]} (g - g'(w_i))^2 k_g(w_i | v_i) \bar{q}(dg).$$

Eine Minimisierung von $\bar{r}(\bar{q}, g')$ liefert bei Verwendung von S2.23 Bedingung 3) unter der Voraussetzung

$$\int_{[0,1]} k_g(w_i | v_i) \bar{q}(dg) > 0$$

die eindeutige Bayes-EV $\varepsilon_{\bar{q}}$ durch die Schätzung

$$g'_{\bar{q}}(w_i) = \frac{\int_{[0,1]} g k_g(w_i | v_i) \bar{q}(dg)}{\int_{[0,1]} k_g(w_i | v_i) \bar{q}(dg)}$$

und das dazugehörige mittlere Risiko

$$\bar{r}(\bar{q}, g'_{\bar{q}}) = \int_{[0,1]} g^2 \bar{q}(dg) - \sum_{v^1} \dots \sum_{v^n} \frac{[\int_{[0,1]} g \prod_{i=1}^n k_g(v^i | u) \bar{q}(dg)]^2}{\int_{[0,1]} \prod_{i=1}^n k_g(v^i | u) \bar{q}(dg)}.$$

Man erkennt sofort, daß es nicht einfach ist, $\bar{r}(\bar{q}, g_{\bar{q}}^1)$ als Funktion von $\bar{q}$ zu maximieren, um eine am wenigsten günstige a priori Verteilung zu bestimmen und weiter durch die entsprechende eindeutig festgelegten Bayes-EV eine Minimax-EV zu erhalten. Im vorliegenden Fall könnte man auch eine Minimax-EV finden, indem man eine EV ε_0 angibt, für die $r(k_g, g_0')$ nicht von g abhängt und nachweist, daß ε_0 eine Bayes-EV ist. Auf diese Methode sei hier nicht weiter eingegangen (siehe [18], [15]). Es ist ebenfalls klar, daß man sie hier nicht ohne Mühen anwenden kann.

Betrachtet man $\bar{r}(\bar{q}, g_{\bar{q}}')$ als Funktion von $\bar{q}$, dann fragt man nach dem asymptotischen Verhalten dieser Funktion. Dazu sei folgendes bemerkt:

Wenn $k_1(\cdot|u) = k_2(\cdot|u)$, dann ist

$$\bar{r}(\bar{q}, g_{\bar{q}}') = \int_{[0,1]} g^2 \bar{q}(dg) - \left[\int_{[0,1]} g \bar{q}(dg) \right]^2 = \text{const.}$$

Andernfalls gibt es ein v , so daß $k_1(v|u) \neq k_2(v|u)$.

Ein solches v sei festgelegt. Eine naheliegende Schätzung für g ist dann

$$g_v'(\omega_1) = \frac{1}{k_1(v|u) - k_2(v|u)} [h_v(\omega_1) - k_2(v|u)],$$

wobei $h_v(\omega_i)$ die relative Häufigkeit dafür ist, daß $v^i = v$ ($i = 1, \dots, n$). Wegen der statistischen Unabhängigkeit der Reaktionen der Kopien gilt

$$\mathbb{E}_{k_g} h_v(\omega_1) = k_g(v|u)$$

$$\mathbb{E}_{k_g} h_v^2(\omega_1) = \frac{1}{n} k_g(v|u) [1 - k_g(v|u)] + k_g^2(v|u).$$

Damit ist g_v' erwartungstreu, d.h.

$$\mathbb{E}_{k_g} g_v'(\omega_1) = g$$

und es gilt

$$\bar{r}(k_g, g_v') = \mathbb{E}_{k_g} (g - g_v')^2 = \frac{1}{n} \left[\frac{k_2(v|u) [1 - k_2(v|u)]}{[k_1(v|u) - k_2(v|u)]^2} + \frac{1 - 2k_2(v|u)}{k_1(v|u) - k_2(v|u)} g - g^2 \right].$$

Also ist auch $\bar{r}(\bar{q}, g'_v) = O(\frac{1}{n})$. Da aber $\bar{r}(\bar{q}, g'_q) \leq \bar{r}(\bar{q}, g'_v)$, folgt, daß $\bar{r}(\bar{q}, g'_q)$ mindestens wie $\frac{1}{n}$ mit $n \rightarrow \infty$ gegen Null konvergiert und $g'_q(\omega_i)$ im quadratischen Mittel gegen die nach $\bar{q}$ verteilte Zufallsvariable g konvergiert.

Den Abschätzungen des mittleren Risikos $\bar{r}$ und der Untersuchung des asymptotischen Verhaltens dieser Funktion in Abhängigkeit von der Zahl n der Kopien wird in Kapitel 3 für zustandsbestimmende Vorgabe-EV die Aufmerksamkeit gewidmet werden. Für allgemeine adaptive EV dürften solche Betrachtungen sehr kompliziert sein.

- 3) Nichtadaptive Experimente fanden bisher in der Theorie der Experimente mit deterministischen Automaten große Beachtung. Insbesondere waren dies diagnostische und heimführende Experimente. Es war z.B. ein kürzestes Eingabewort gesucht, das ein einfaches heimführendes Experiment gestattete. Das "gestattete" bedeutet in der Terminologie dieser Arbeit, daß der maximale bedingte Entscheidungsfehler bei Benutzung von u gleich Null ist. Die Kostenfunktion kann man dann proportional zur Länge der Eingabeworte wählen. Insbesondere galt die Aufmerksamkeit häufig Abschätzungen über die Länge eines kürzesten Experimentes. Es sind solche Dinge, abgesehen von den zahlreichen Originalarbeiten in dem Buch von Gill [11], abgehandelt. Für einfache diagnostische Vorgabe-Experimente mit stochastischen Automaten findet man dazu etwas in [4] und etwas allgemeiner in Kapitel 3 dieser Arbeit. Auf dem Gebiet der Mustererkennung spielt die zu dem Problem, günstige nichtadaptive EV zu finden, analoge Aufgabe, günstige Klassifikationsvorschriften zu finden, die mittlere Klassifikationsverluste plus mittlere Klassifikationskosten minimieren, ebenfalls eine gewisse Rolle (vergleiche z.B. [5]).

- 4) Ausgabenvorhersage- und ähnliche Probleme mit stochastischen Automaten sind verwandt mit einfachen zustandsbestimmenden Vorgabeexperimenten. Deshalb sei die Arbeit von Booth [6] in diesem Zusammenhang erwähnt.
- 5) Carlyle (vergl. [7]) hat gezeigt, daß das Eingabe-Ausgabe-Verhalten von stochastischen Automaten mit Anfangsverteilung vollständig durch die bedingte Wahrscheinlichkeiten für alle Eingaben bis zur Länge $2|Z|-1$, wenn Z die Menge der Zustände ist, beschrieben werden kann. Es sind nämlich die Wahrscheinlichkeiten für Eingabewörter größerer Länge aus denen für Eingabewörter bis zur Länge $2|Z|-1$ berechenbar. Daraus folgt, daß man bei Experimenten mit Automaten von den entsprechenden Kanälen auch nur eine endliche Zahl von W_k $k(1 \leq k \leq |Z|)$ zu wissen braucht, um den Kanal vollständig beschreiben zu können. Welche Vereinfachungen dieser Sachverhalt auf die Konstruktion von EV mit sich bringen könnte, ist noch nicht bekannt. Weiterhin zeigt Carlyle in diesem Artikel, unter welchen Bedingungen man prinzipiell aus dem Eingabe-Ausgabe-Verhalten bis auf Isomorphie eindeutig auf die innere Struktur des Automaten schließen kann. Diese Bedingungen sind also äußerst wichtig bei Automatenidentifikationsexperimenten.
- 6) Für zustandsbestimmende Experimente ist folgendes zu beachten: Zwei Zustände sollten allein mit Hilfe von Ein- und Ausgaben unterscheidbar sein. Ist das nicht der Fall, so sind die Zustände äquivalent. Es ist also zweckmäßig, die Automaten als reduziert vorauszusetzen, d.h. sie besitzen keine äquivalenten Zustände mehr. Zustandsreduktionsprobleme werden z.B. in den Arbeiten [8] und [16] behandelt. Äquivalente Automaten, d.h. also solche, die sich in ihrem Ein-Ausgabe-Verhalten nicht unterscheiden, sind durch Experimente nicht unterscheidbar. Es ist dem-

nach sinnvoll, eine Äquivalenzklasseneinteilung der Automaten vorzunehmen und jeweils nur Repräsentanten verschiedener Äquivalenzklassen durch Experimente unterscheiden zu wollen.

3. Einige Aussagen über zustandsbestimmende Vorgabeexperimente mit stochastischen Automaten

In Kapitel 2 wurde eine allgemeine Theorie der Experimente mit Kanälen ausgearbeitet, die auch eine allgemeine Theorie der Experimente mit stochastischen Automaten darstellt. Die Gedanken aus 2.5 weiter verfolgend wird es durch Präzisierung der Aufgabenstellung eines Experimentes möglich, zusätzliche Aussagen über spezielle Experimente zu gewinnen. In diesem Kapitel werden Vorgabeexperimente mit Automaten behandelt, wie sie in 2.3 Fall 1) vorgestellt wurden. Wenn Z die Menge der Zustände des Automaten ist, sei also $K = \{k_z : z \in Z\}$ und zusätzlich $A = Z$. Die dadurch charakterisierten Experimente heißen zustandsbestimmend. Die Funktionale k_z werden aus dem Funktional p und ggf. aus der Anfangsverteilung q des Automaten bestimmt. Die Verlustfunktion sei $v(k_z, z') = 1 - \delta_{zz'}$ ⁺⁾ . Man interessiert sich demnach für erwartete Fehler, und alle Aussagen aus 2.3 ab p. 43 sind gültig. Das Hauptinteresse des vorliegenden Kapitels gilt den Abschätzungen der erwarteten Fehler. Es werden dabei Abschätztechniken benutzt, die schon in der Informationstheorie angewandt wurden. In diesem Zusammenhang sind z.B. die Arbeiten von Gallager und Augustin zu nennen. Man findet dazu etwas in den Büchern von Gallager [10] und Wolfowitz [19] und in den Originalarbeiten von Shannon, Gallager, Berlekamp [17] sowie denen von Augustin [1], [2].

⁺⁾ $\delta_{zz'} := \begin{cases} 1, & \text{wenn } z = z' \\ 0 & \text{sonst} \end{cases}$

3.1 Schranken für den Entscheidungsfehler bei n-fach diagnostischen Vorgabeexperimenten

Ein Automat werde wie in 1.1 beschrieben durch (X, Y, Z, p) . Sei für $u \in X^*$ und $v \in Y^*$ mit $l(v) = l(u)$

$$k_z(v|u) := p^z(v|u) := \sum_{z \in Z} p(v, z' | z, u)$$

und wenn $v_1 = (u^1, \dots, u^n)$, $w_1 = (v^1, \dots, v^n)$ mit $\ell(w) = \ell(v)$ ⁺⁾ , dann die schon gewohnte Verallgemeinerung

$$k_z(w|v) := p^z(w|v) := \prod_{i=1}^n p^z(v^i | u^i).$$

Also kann $p^z(w|v)$ als die w_k dafür gedeutet werden, daß für $i=1, \dots, n$ die i -te Kopie des Automaten v^i ausgibt, wenn der Anfangszustand z war und u^i eingegeben wurde. Eine Anfangsverteilung q über Z spielt die Rolle einer a priori Verteilung $\bar{q} = q$. Die solcherart festgelegten zustandsbestimmenden n-fach Vorgabeexperimente heißen diagnostische n-fach Vorgabeexperimente. Mit ihnen bestimmt der ER also den ihm unbekannten Anfangszustand des ihm in seiner Struktur bekannten Automaten. Dabei wird natürlich vorausgesetzt, daß der Automat mindestens zwei Zustände besitzt. Die zugehörigen EV sollen mit ε_d bezeichnet werden.

Nach den abschließenden Bemerkungen aus 2.5 ist es sinnvoll vorauszusetzen, daß der Automat voll reduziert ist. D.h., es gibt keine zwei verschiedenen Zustände z, z' aus Z mit der Eigenschaft, daß für alle $u \in X^*$ $p^z(\cdot | u) = p^{z'}(\cdot | u)$. Andernfalls wären zwei Zustände nicht durch ein Experiment unterscheidbar.

Der bedingte Entscheidungsfehler (2.48) wird jetzt bei Eingabe von u (etwas ungenau) mit $r(z, \varepsilon_d(\cdot | u))$ bezeichnet, und der mittlere Entscheidungsfehler (2.49) ist $\bar{r}(q, \varepsilon_d(\cdot | v_1, \cdot))$.

⁺⁾ Der Index 1 wird im folgenden weggelassen.

Die minimalen Entscheidungsfehler seien

$$(3.1) \quad p_{fd}(v) := \min_{\varepsilon_d} \max_{z \in Z} r(z, \varepsilon_d(\cdot | v, \cdot)) ,$$

$$(3.2) \quad \bar{p}_{fd}(v) := \min_{\varepsilon_d} \bar{r}(q, \varepsilon_d(\cdot | v, \cdot)) = 1 - \sum_{\omega} \max_{z \in Z} q(z) p^2(\omega | v).$$

3.1.1 Obere Schranken

Aus (3.2) folgt offensichtlich die Abschätzung

$$(3.3) \quad \bar{p}_{fd}(q, v) \leq \sum_z \sum_{\omega} q(z) p^2(\omega | v) \cdot \left[1 - \prod_{z' \neq z} 1_{\{q(z) p^2(\cdot | v) > q(z') p^2(\cdot | v)\}}(\omega) \right]^{+1}$$

Diese Abschätzung ist i.a. nicht verbesserbar. Das Gleichheitszeichen gilt genau dann, wenn eine Bayes-EV bzgl. q bis auf Nullmengen von Elementen ω bzgl. des Maßes

$\sum_z q(z) p^2(\cdot | v)$ eindeutig festgelegt ist. Die Berechnung der rechten Seite von (3.3) setzt die volle Kenntnis der Maße q und $p^2(\cdot | v)$ voraus. Das Ziel ist es, Abschätzungen nach oben für $p_{fd}(q, v)$ zu finden, die weniger Kenntnisse über die zugrunde liegenden Maße benötigen. Es wird sich zeigen, daß man die rechte Seite von (3.3) weiter in Richtung des angegebenen Zieles abschätzen kann und zudem Aussagen über die Abhängigkeit der Größe $p_{fd}(q, v)$ von der Zahl n der Kopien erhält.

Sei s reell und ≥ 0 , dann gilt: Ist [...] in (3.3) für ein ω gleich 0, dann ist

$$\left[\frac{\sum_{z' \neq z} q(z') p^2(\omega | v)}{q(z) p^2(\omega | v)} \right]^s \geq 0 \quad \text{und sonst } \geq 1. \text{ Dann}$$

⁺) $1_B(b) := \begin{cases} 1, & \text{wenn } b \in B \\ 0 & \text{sonst} \end{cases}, \quad \{f_1(\cdot) > f_2(\cdot)\} := \{b: f_1(b) > f_2(b)\}$

folgt aus (3.3)

$$\bar{P}_{fd}(q, v) \leq \sum_z \sum_w q(z) p^z(w|v) \left[\frac{\sum_{z' \neq z} q(z') p^{z'}(w|v)}{q(z) p^z(w|v)} \right]^s$$

Schränkt man s auf Werte ≤ 1 ein, so kann man weiter abschätzen

$$(3.4) \quad \sum_z \sum_w q(z)^{1-s} p^z(w|v)^{1-s} \left[\sum_{z' \neq z} q(z') p^{z'}(w|v) \right]^s \\ \leq \sum_z \sum_w q(z)^{1-s} p^z(w|v)^{1-s} \sum_{z' \neq z} q(z')^s p^{z'}(w|v)^s =: h(s),$$

wobei $h(1) = h(0) = |Z| - 1$ ist und Summanden, für die $q(z) p^z(w|v) = 0$ ist, fortzulassen sind. Im Bereich $0 < s < 1$ ist dann $h(s)$ eine zweimal stetig differenzierbare Funktion von s mit nichtnegativer zweiter Ableitung. Diese ist gleich Null genau dann, wenn $q(z) p^z(w|v)$ eine Gleichverteilung auf der Menge der Paare (z, w) definiert. (Dann ist aber $h(s) = 1$ ($0 < s < 1$).) Also ist h eine konvexe Funktion von s in $]0, 1[$. Aus (3.4) erkennt man durch Umordnen der Summen, daß $h(s) = h(1-s)$. Damit nimmt h in $]0, 1[$ sein Minimum bei $s = \frac{1}{2}$ an. Indem man die ursprüngliche Bedeutung für $p^z(w|v)$ einsetzt und umordnet, erhält man den folgenden Satz.

S3.1 Satz: Der minimale mittlere Entscheidungsfehler bei Vorlage der Anfangsverteilung q und der durch $v = (u^1, \dots, u^n)$ festgelegten Eingaben genügt der Ungleichung

$$(3.5) \quad \bar{P}_{fd}(q, v) \leq \sum_z \sum_{z' \neq z} q(z)^{\frac{1}{2}} q(z')^{\frac{1}{2}} \prod_{i=1}^n \sum_v p^z(v|u^i)^{\frac{1}{2}} p^{z'}(v|u^i)^{\frac{1}{2}}.$$

Man fragt sofort nach dem Verhalten der rechten Seite von (3.5) für wachsende n . Dazu sei bemerkt:

$\sum_v p^z(v|u)^{\frac{1}{2}} p^{z'}(v|u)^{\frac{1}{2}}$ ist genau dann kleiner als 1, wenn $p^z(\cdot|u) \neq p^{z'}(\cdot|u)$ (vergl. z.B. [10] p. 522).

Man stelle sich w als Anfangsstück einer Folge $(u^i)_{i \in \mathbb{N}}$ von Eingabeworten u^i vor. Dann liefert ein bekannter Satz über die Nullkonvergenz unendlicher Produkte (siehe z.B. [13]) das

K3.2 Korollar: Wenn es für je zwei verschiedene z, z' aus Z eine Teilfolge $(u_j^{z,z'})_{j \in \mathbb{N}}$ der Folge $(u^i)_{i \in \mathbb{N}}$ von Eingaben u^i in die Kopien mit der Nummer i gibt mit den Eigenschaften

$$1) \text{ Für alle } j \in \mathbb{N} \text{ ist } p^z(\cdot | u_j^{z,z'}) \neq p^{z'}(\cdot | u_j^{z,z'}),$$

$$2) \sum_{j=1}^n \left[1 - \left[\sum_{\nu} p^z(\nu | u_j^{z,z'})^{\frac{1}{2}} p^{z'}(\nu | u_j^{z,z'})^{\frac{1}{2}} \right]^2 \right] \xrightarrow{n \rightarrow \infty} \infty,$$

dann konvergiert $\bar{p}_{fd}(q, \nu)$ für jedes q und $p_{fd}(\nu)$ mit $n \rightarrow \infty$ gegen Null.

Daß die Behauptung auch für $p_{fd}(\nu)$ gültig ist, folgt aus der Existenz einer am wenigsten günstigen a priori Verteilung (also Anfangsverteilung) q' mit der Eigenschaft $\max_q \bar{p}_{fd}(q, \nu) = \bar{p}_{fd}(q', \nu) = p_{fd}(\nu)$.

Da der vorliegende Automat vollreduziert ist, sind die Voraussetzungen von K3.2 leicht erfüllbar. Zu je zwei ungleichen Zuständen z_1 und z_2 gibt es ein Wort $u^{z_1, z_2} \in X^*$ das z_1 und z_2 trennt, d.h. $p^{z_1}(\cdot | u^{z_1, z_2}) \neq p^{z_2}(\cdot | u^{z_1, z_2})$.

Es gibt nach [8] sogar schon solche Worte u^{z_1, z_2} mit der Länge $l(u^{z_1, z_2}) \leq |Z| - 1$. Um alle Zustände voneinander trennen zu können, benötigt man also $|Z|(|Z| - 1)/2$ dieser Wörter u^{z_1, z_2} , gerade so viele, wie es verschieden Zweiermengen $\{z_1, z_2\}$ in Z gibt. (Dabei müssen die u^{z_1, z_2} nicht notwendig alle paarweise verschieden sein.)

Sei w' ein $|Z|(|Z| - 1)/2$ -tupel der Wörter u^{z_1, z_2} . Dann erfüllt die Folge, die durch periodische Wieder-

holung von w' entsteht, die Voraussetzungen von K3.2. Bei n' -facher Wiederholung von w' gilt

$$(3.6) \quad \bar{p}_{fd}(q, w) \leq \sum_z \sum_{z' \neq z} q(z)^{\frac{1}{2}} q(z')^{\frac{1}{2}} \exp [-n' \bar{c}_o^{zz'}(w')]$$

mit

$$(3.7) \quad \bar{c}_o^{zz'} := - \sum_{\{z_1, z_2\}} \ln \sum_v p^z(v|u^{z_1, z_2})^{\frac{1}{2}} p^{z'}(v|u^{z_1, z_2})^{\frac{1}{2}} > 0,$$

wobei in (3.7) über alle Zweiermengen $\{z_1, z_2\} \subset Z$ summiert wird.

Hat man irgend ein $|Z|(|Z|-1)/2$ -tupel w^* von Eingabewörtern, so ist die Bedingung $\bar{c}_o^{zz'}(w^*) > 0$ für alle Zweiermengen $\{z, z'\}$ aus Z hinreichend dafür, daß für die durch periodische Wiederholung von w^* entstehende Folge die Aussagen aus K3.2 gelten.

Möchte man bei einem diagnostischen Vorgabeexperiment den maximalen bedingten Entscheidungsfehler $p_{fd}(w)$ unter eine Schranke $p_o > 0$ drücken, so folgt aus (3.6) eine grobe Abschätzung für die Zahl $n = n' |Z|(|Z|-1)/2$ der benötigten Kopien.

K3.3 Korollar: Wenn w' aus $|Z|(|Z|-1)/2$ Wörtern besteht, die insgesamt alle Zustände untereinander trennen und die Folge w von Eingaben u^i in die Kopien i ($i=1, \dots, n$) durch periodische Wiederholung der Wörter aus w' entsteht, dann reichen

$$n \geq \frac{|Z|(|Z|-1)}{2 \bar{c}_o(w')} \ln \frac{|Z|-1}{p_o}$$

mit

$$\bar{c}_o(w') := \min_{\{z, z'\}} \bar{c}_o^{zz'}(w')$$

Kopien zum Experimentieren aus, damit für alle q gilt

$$\bar{p}_{fd}(q, w) \leq p_{fd}(w) \leq p_o.$$

Für kleine n , insbesondere für $n=1$ (dann heißt das Experiment einfach), wird die Schranke in (3.5) i.a. nicht sehr genau sein. Es kann sogar vorkommen, daß die rechte Seite dieser Ungleichung größer als 1 ist. Dann muß man sich mit der aus (3.2) unmittelbar folgenden Schranke

$$\bar{p}_{fd}(q, v) \leq 1 - \max_{z \in Z} q(z)$$

zufrieden geben. Diese Schranke ist in manchen Fällen nicht zu verbessern, z.B. dann, wenn $\max_z q(z) p^z(\omega|v)$ für ein festes z_0 und alle ω gleich $q(z_0) p^{z_0}(\omega|v)$ ist.

Für den Spezialfall, daß q die Gleichverteilung q^0 der Anfangszustände ist, d.h. $q^0(z) = 1/|Z|$ für $z \in Z$, läßt sich die Abschätzung (3.5) noch verschärfen. Von (3.3) ausgehend erhält man

$$\begin{aligned} \bar{p}_{fd}(q^0, v) &\leq \frac{1}{|Z|} \sum_z \sum_{\omega} p^z(\omega|v) 1_{\{p^z(\cdot|v) \leq \max_{z' \in Z} p^{z'}(\cdot|v)\}}(\omega) \\ &\leq \frac{1}{|Z|} \sum_z \sum_{z' \in Z} \sum_{\omega} p^z(\omega|v) 1_{\{p^z(\cdot|v) \leq p^{z'}(\cdot|v)\}}(\omega). \end{aligned}$$

Offensichtlich gilt für $0 \leq s \leq 1$

$$\sum_{\omega} p^z(\omega|v) 1_{\{p^z(\cdot|v) \leq p^{z'}(\cdot|v)\}} \leq \sum_{\omega} p^z(\omega|v) \left[\frac{p^{z'}(\omega|v)}{p^z(\omega|v)} \right]^s,$$

wobei alle Summanden der rechten Seite der Ungleichung fortzulassen sind, bei denen entweder $p^z(\omega|v)$ oder $p^{z'}(\omega|v)$ gleich 0 ist. Die rechte Seite der Ungleichung ist konvex in $0 \leq s \leq 1$ und stetig in $0 < s < 1$, sie nimmt also ihr Minimum an einer Stelle $s = s_{zz'} = s_{zz'}(v)$ mit $0 < s_{zz'} < 1$ an. Dann folgt unter Benutzung der ursprünglichen Bedeutungen S3.4.

S3.4 Satz: Der minimale Entscheidungsfehler bei Vorlage Gleichverteilung q^0 der Anfangszustände und der durch v festgelegten Eingaben genügt der Ungleichung

$$(3.8) \quad \bar{p}_{fd}(q^0, v) \leq \frac{1}{|Z|} \sum_z \sum_{z' \in Z} \prod_{i=1}^n \sum_v p^z(v|u_i)^{1-s_{zz'}(u_i)} p^{z'}(v|u_i)^{s_{zz'}(u_i)}.$$

Daß diese Abschätzung für den Fall $q=q^0$ i.a. eine Verschärfung von (3.5) darstellt, folgt aus der Eigenschaft

$$\prod_{i=1}^n \sum_v p^z(v|u^i)^{1-s_{zz'}} p^{z'}(v|u^i)^{s_{zz'}} \leq \prod_{i=1}^n \sum_v p^z(v|u^i)^{\frac{1}{2}} p^{z'}(v|u^i)^{\frac{1}{2}}.$$

Das Gleichheitszeichen gilt höchstens, wenn $p^z(\cdot|v) = p^{z'}(\cdot|v)$ oder wenn $s_{zz'}(v) = \frac{1}{2}$ ist.

Die Voraussetzungen, die zu (3.6) führten, liefern im Falle $q=q^0$ folgendes: Es ist dann

$$\sum_{\omega} p^z(\omega|v)^{1-s} p^{z'}(\omega|v)^s = \left[\sum_{\omega'} p^z(\omega'|v')^{1-s} p^{z'}(\omega'|v')^s \right]^{n'}.$$

Man sieht mittels Differentiation, daß dieser Ausdruck durch $s=s_{zz'}(v')$ minimisiert wird. Also erhält man aus (3.6)

$$(3.9) \quad \bar{p}_{fd}(q^0, v) \leq \sum_z \sum_{z' \neq z} \exp \left[-\ln |Z| - n' c_0^{zz'}(v') \right],$$

wobei

$$(3.10) \quad c_0^{zz'}(v') = - \sum_{(z_1, z_2)} \ln \sum_v p^z(v|u^{z_1, z_2})^{1-s_{zz'}(v')} p^{z'}(v|u^{z_1, z_2})^{s_{zz'}(v')}$$

mit der Eigenschaft

$$c_0^{zz'}(v') \geq \bar{c}_0^{zz'}(v').$$

3.1.2 Untere Schranken für große n

Untere Schranken für $\bar{p}_{fd}(q, v)$ zu finden, aus denen man das asymptotische Verhalten dieses Fehlers ablesen kann, ist im allgemeinen Fall nicht mehr einfach. Die im vorliegenden Abschnitt angegebenen Schätzmethoden liefern nur Schranken für $\bar{p}_{fd}(q^0, v)$, wobei q^0 die Gleichverteilung der Anfangszustände ist. Zunächst wird eine Abschätzung mit Hilfe der Tschebyscheff-Ungleichung erhalten und anschließend eine zweite unter Benutzung des

Normalapproximationstheorems von Berry-Eséeen, die für hinreichend große n eine Verbesserung der ersten Abschätzung darstellt.

Beschreibe $u = (u^1, \dots, u^n)$ wieder die Eingabeworte in die Kopien i ($i = 1, \dots, n$). Sei D die Menge der möglichen Ausgaben w mit $\ell(w) = \ell(u)$, wenn u eingegeben wird. Eine Bayes-EV $\mathcal{E}'_d$ kann auch durch die Vorschrift:

Für jedes w gibt es ein z , so daß $\mathcal{E}_d(z|u, w) = 1$ beschrieben werden. $\mathcal{E}'_d$ operiert also im wesentlichen deterministisch. In diesem Abschnitt werden deshalb nur noch deterministisch operierende EV $\mathcal{E}_d$ betrachtet. Dann gibt es eine direkte Zerlegung $\{D_z: z \in Z\}$ von D , so daß $w \in D_z$ genau dann, wenn $\mathcal{E}_d(z|u, w) = 1$, also $\bigcup_{z \in Z} D_z = D$ und $D_z \cap D_{z'} = \emptyset$, wenn $z \neq z'$. Sei $D'_z := D - D_z$.

Betrachtet werde ein festes $\mathcal{E}_d$ und die zugehörige direkte Zerlegung $\{D_z\}$. Seien t_1, t_2 und s reelle Zahlen mit $0 < t_1 < t_2$ und $0 \leq s \leq 1$. Zwei Zustände $z \neq z'$ seien festgelegt. Dann kann man mit der offensichtlichen Bezeichnung

$p^z(D'|u)$ für $D' \subset D$ schätzen:

$$(3.11) \quad p^z(D'_z|u) \geq p^z(D_z \cap \left\{ \frac{p^z(\cdot|u)}{p^{z'}(\cdot|u)} > t_1 \right\} | u) > \\ > t_1^s \sum_w p^z(w|u) \left[\frac{p^z(w|u)}{p^{z'}(w|u)} \right]^{-s} 1_{D_z \cap \left\{ \frac{p^z(\cdot|u)}{p^{z'}(\cdot|u)} > t_1 \right\}}(w),$$

$$(3.12) \quad p^{z'}(D_z|u) \geq p^{z'}(D_z \cap \left\{ \frac{p^z(\cdot|u)}{p^{z'}(\cdot|u)} < t_2 \right\} | u) > \\ > t_2^{-(1-s)} \sum_w p^{z'}(w|u) \left[\frac{p^z(w|u)}{p^{z'}(w|u)} \right]^{1-s} 1_{D_z \cap \left\{ \frac{p^z(\cdot|u)}{p^{z'}(\cdot|u)} < t_2 \right\}}(w) = \\ = t_2^{-(1-s)} \sum_w p^z(w|u) \left[\frac{p^z(w|u)}{p^{z'}(w|u)} \right]^{-s} 1_{D_z \cap \left\{ \frac{p^z(\cdot|u)}{p^{z'}(\cdot|u)} < t_2 \right\}}(w).$$

Multipliziert man (3.11) mit t_1^{-s} , (3.12) mit t_2^{1-s} , addiert man die Ungleichungen, verkleinert die Mengen

$$\left\{ \frac{p^z(\cdot|v)}{p^{z'}(\cdot|v)} > t_1 \right\} \quad \text{und} \quad \left\{ \frac{p^z(\cdot|v)}{p^{z'}(\cdot|v)} < t_2 \right\} \quad \text{zu} \\ \left\{ t_1 < \frac{p^z(\cdot|v)}{p^{z'}(\cdot|v)} < t_2 \right\} \quad \text{und setzt schließlich } t_i' := \ln t_i \quad (i=1,2),$$

dann erhält man

$$(3.13) \quad e^{-st_1'} p^z(D_z|v) + e^{(1-s)t_2'} p^{z'}(D_z|v) \\ \geq \sum_w p^z(w|v)^{1-s} p^{z'}(w|v)^s \mathbb{1}_{\left\{ t_1' < \frac{p^z(\cdot|v)}{p^{z'}(\cdot|v)} < t_2' \right\}}(w).$$

Es ist wieder zu beachten, daß alle Summanden, für die $p^z(w|v) = 0$ oder $p^{z'}(w|v) = 0$ ist, fortzulassen sind.

Für die weiteren Untersuchungen wird die folgende Voraussetzung benötigt:

V3.1 Sei $v = (u^1, \dots, u^n)$. Dann gibt es eine Zahl $b_0(v) > 0$, so daß für alle $i, z, z' \neq z$ und $0 < s < 1$

$$\sum_v p^z(v|u^i)^{1-s} p^{z'}(v|u^i)^s \geq b_0(v).$$

Der Ausdruck

$$\sum_w p^z(w|v)^{1-s} p^{z'}(w|v)^s = \prod_{i=1}^n \sum_v p^z(v|u^i)^{1-s} p^{z'}(v|u^i)^s$$

wird nach den Bemerkungen, die zu (3.8) führten, durch $s = s_{zz'} = s_{zz'}(v)$ mit $0 < s_{zz'} < 1$ minimiert. Sei

$$(3.14) \quad b_0^{zz'}(u^i) := \sum_v p^z(v|u^i)^{1-s_{zz'}(v)} p^{z'}(v|u^i)^{s_{zz'}(v)}.$$

Dann ist mit V3.1

$$(3.15) \quad p^{zz'}(\cdot|v) := p^z(\cdot|v)^{1-s_{zz'}(v)} p^{z'}(\cdot|v)^{s_{zz'}(v)} \prod_{i=1}^n b_0^{zz'}(u^i)^{-1}$$

ein Wk-maß auf D . Man erhält durch einfaches Nachrechnen

$$\mathbb{E}_{p^{zz'}(\cdot|v)} \ln \frac{p^z(\cdot|v)}{p^{z'}(\cdot|v)} = \prod_{i=1}^n b_0^{zz'}(u^i)^{-1} \frac{d}{ds} \sum_w p^z(w|v)^{1-s} p^{z'}(w|v)^s \Big|_{s=s_{zz'}}$$

Der unter $\frac{d}{ds}$ stehende Ausdruck ist differenzierbar und nimmt sein Minimum bei $s=s_{zz'}$ an, daher ist notwendig

$$\mathbb{E}_{p^{zz'}(\cdot|v)} \ln \frac{p^z(\cdot|v)}{p^{z'}(\cdot|v)} = 0.$$

Setzt man in (3.13) $s=s_{zz'}$, dann ist es sinnvoll, t_1' und t_2' symmetrisch um den Nullpunkt anzusetzen, wenn also $t' > 0$ ist, dann $t_1' = -t'$ und $t_2' = t'$. Die linke Seite von (3.13) wird dann vergrößert, indem man für $e^{-st_1'}$ und für $e^{-(1-s)t_2'}$ die Zahl $e^{t'}$ setzt. Darauf ergibt sich mit (3.14) und (3.15)

$$(3.16) \quad p^z(D_z|v) + p^{z'}(D_z|v) \geq e^{-t'} \prod_{i=1}^n b_o^{zz'}(u^i) \mathbb{E}_{p^{zz'}(\cdot|v)} 1_{\{-t' < \ln \frac{p^z(\cdot|v)}{p^{z'}(\cdot|v)} < t'\}}.$$

Sei

$$(3.17) \quad t' := \left[2 \mathbb{E}_{p^{zz'}(\cdot|v)} \left[\ln \frac{p^z(\cdot|v)}{p^{z'}(\cdot|v)} \right]^2 \right]^{\frac{1}{2}},$$

dann ist es möglich, die Tschebyscheff-Ungleichung auf den Erwartungswert in (3.16) anzuwenden, und man erhält

$$(3.18) \quad p^z(D_z|v) + p^{z'}(D_z|v) \geq \frac{1}{2} \prod_{i=1}^n b_o^{zz'}(u^i) \cdot \exp \left\{ - \left[2 \mathbb{E}_{p^{zz'}(\cdot|v)} \left[\ln \frac{p^z(\cdot|v)}{p^{z'}(\cdot|v)} \right]^2 \right]^{\frac{1}{2}} \right\}.$$

Zunächst muß die Frage geklärt werden, ob die rechte Seite von (3.17) existiert. Sei für $j \in \mathbb{N}$

$$(3.19) \quad b_j^{zz'}(u^i) := b_o^{zz'}(u^i)^{-1} \sum_v p^z(v|u^i)^{1-s_{zz'}} p^{z'}(v|u^i)^{s_{zz'}} \left| \ln \frac{p^z(v|u^i)}{p^{z'}(v|u^i)} \right|^j.$$

Betrachtet man zunächst die Menge aller v , für die

$$p^{z'}(v|u^i) / p^z(v|u^i) \leq 1 \quad \text{und dann diejenige, für die}$$

$$p^z(v|u^i) / p^{z'}(v|u^i) \leq 1, \quad \text{dann führt die folgende Überlegung zum Ziel: die Funktion } g | \ln g|^j,$$

wenn man ihr den Wert 0 bei $g=0$ gibt, ist stetig und nicht größer als $j^j e^{-j}$ im Bereich $0 \leq g \leq 1$. Dann findet man nach einer

kleinen Umformung der rechten Seite von (3.19) unter Beachtung von V3.1 und $0 < s_{zz'} < 1$

$$b_j^{zz'}(u^i) \leq s_{zz'}^{-j} (1 - s_{zz'})^{-j} e^{-j} b_0(v)^{-1} < \infty.$$

Damit ist für $j \in \mathbb{N}$

$$\mathbb{E}_{p^{zz'}(\cdot|v)} \left| \ln \frac{p^z(\cdot|v)}{p^{z'}(\cdot|v)} \right|^j = \sum_{i=1}^n b_j^{zz'}(u^i) < \infty.$$

Führt man auf (3.18) die Operation $\frac{1}{|Z|(|Z|-1)} \sum_z \sum_{z' \neq z} \dots$

aus, so ergibt sich

$$\begin{aligned} (3.20) \quad \bar{r}(q^0, \varepsilon_d(\cdot|v, \cdot)) &= \frac{1}{|Z|} \sum_z p^z(cD_z|v) \\ &\geq \frac{1}{2|Z|^2} \sum_z \sum_{z' \neq z} \prod_{i=1}^n b_0^{zz'}(u^i) \cdot \exp \left[-2 \sum_{i=1}^n b_2^{zz'}(u^i) \right]^{\frac{1}{2}}. \end{aligned}$$

Sei

$$b_2(v) := \max_{i, z \neq z'} \left[2 b_2^{zz'}(u^i) \right]^{\frac{1}{2}},$$

dann ist der folgende Satz bewiesen.

S3.5 Satz: Beschreibe $v = (u^1, \dots, u^n)$ die Menge der Eingabeworte u^i in die Kopien i ($i=1, \dots, n$). Die Bedingung V3.1 sei erfüllt. q^0 bezeichne die Gleichverteilung der Anfangszustände. Dann gibt es nur von v und dem Automaten abhängige Zahlen $0 < b_2(v) < \infty$ und $0 < s_{zz'}(v) < 1$, so daß

$$(3.21) \quad P_{fd}(v) \geq \bar{P}_{fd}(q^0, v) \geq \exp \left[-\ln 2|Z|^2 - \sqrt{n} b_2(v) \right] \cdot \sum_z \sum_{z' \neq z} \prod_{i=1}^n \sum_v p^z(v|u^i)^{1-s_{zz'}(v)} p^{z'}(v|u^i)^{s_{zz'}(v)}.$$

⁺) Da bei der vorliegenden Betrachtung nur deterministisch operierende EV ε_d zugelassen sind, braucht es keine Minimax-EV zu geben. Dann ist unter $P_{fd}(v)$ der Wert $\sup_q \bar{P}_{fd}(q, v)$ zu verstehen, der auch für ein am wenigsten günstiges $q=q'$ von $\bar{P}_{fd}(q, v)$ angenommen wird.

An dieser Stelle ist ein Vergleich von (3.8) und (3.21) angebracht. Beide Abschätzungen enthalten den Term $\sum_z \sum_{z' \neq z} \dots$. Wenn man nun annimmt, daß $\exp[-\sqrt{n'} b_2(v)]$ in Abhängigkeit von n weniger stark fällt als $\sum_z \sum_{z' \neq z} \dots$, dann sind die beiden Abschätzungen asymptotisch scharf. Dieser Fall tritt tatsächlich ein, wie der im folgenden behandelte Spezialfall zeigt.

v' bestehe wie in Abschnitt 3.1.1 aus $|Z|(|Z|-1)/2$ Wörtern u^{z_1, z_2} , die insgesamt alle Zustände des Automaten trennen. v entstehe durch n' -fache periodische Wiederholung von v' . Es sei vorausgesetzt, daß v' die Bedingung V3.1 erfüllt. Offensichtlich ist $b_2(v) = b_2(v')$. Mit den Bemerkungen, die zu (3.9) führten, ist aber auch $s_{22'}(v) = s_{22'}(v')$. Damit erhält man aus (3.21) mit (3.10)

$$(3.22) \quad \bar{p}_{fd}(q^0, v) \geq \sum_z \sum_{z' \neq z} \exp[-\ln 2|Z|^2 - \sqrt{n'} \sqrt{|Z|(|Z|-1)/2} b_2(v') - n' c_0^{zz'}(v')].$$

Aus (3.9) und (3.22) folgt dann, daß sich in dem hier behandelten Spezialfall $\bar{p}_{fd}(q^0, v)$ asymptotisch wie $\sum_z \sum_{z' \neq z} \exp[-n' c_0^{zz'}(v')]$ verhält.

Mit den Definitionen

$$(3.23) \quad c_0(v') := \frac{1}{|Z|(|Z|-1)} \sum_z \sum_{z' \neq z} c_0^{zz'}(v'),$$

$$(3.24) \quad c_2(v') := \frac{1}{|Z|(|Z|-1)} \sum_z \sum_{z' \neq z} \left[2 \sum_{\{z_1, z_2\}} b_2^{zz'}(u^{z_1, z_2}) \right]^{\frac{1}{2}}$$

findet man aus (3.20) unter Anwendung von Jensens Ungleichung

$$(3.25) \quad \bar{p}_{fd}(q^0, v) \geq \exp \left[\ln \frac{|Z|-1}{2|Z|} - \sqrt{n'} c_2(v') - n' c_0(v') \right].$$

Durch die Benutzung der Jensenschen Ungleichung ist die Abschätzung (3.25) zwar für große n weniger scharf als diejenige in (3.22), jedoch liefert sie den Beweis für das folgende Korollar, welches das Gegenstück zu K3.3 darstellt.

K3.6 Korollar: v' bestehe aus $|Z|(|Z|-1)/2$ Wörtern, die insgesamt alle Zustände des Automaten trennen. Die Eingabefolge v entstehe durch n' -fache periodische Wiederholung von v' . v' erfülle die Bedingung V3.1. q^0 sei die Gleichverteilung der Anfangszustände. Wenn p_0 eine reelle Zahl mit

$$0 < p_0 < \frac{|Z|-1}{2|Z|}$$

ist, gilt:

Ist die Zahl der zum Experimentieren benutzten Kopien

$$n \leq \frac{|Z|(|Z|-1)}{8c_0(v')^2} \left[-c_1(v') + \left[c_2(v')^2 - 4c_0(v') \ln \frac{2|Z|p_0}{|Z|-1} \right]^{\frac{1}{2}} \right]^2,$$

so folgt

$$p_{fa}(v) \geq \bar{p}_{fa}(q^0, v) \geq p_0.$$

Die Ungleichung (3.21) kann offensichtlich nur gute Abschätzungen für große n liefern. Für kleine n findet man brauchbare Abschätzungen des mittleren Fehlers nach unten in 3.1.3. Im verbleibenden Teil dieses Abschnittes wird versucht, zumindest für hinreichend große n den Term $\exp \left[-\ln 2 |Z|^2 - \sqrt{n'} b_2(v) \right]$ in (3.21) zu vergrößern.

Ausgehend von (3.16) wird im folgenden $E_{p^{2n}(\cdot|v)} \mathbb{1}_{\{\dots\}}$ mit Hilfe eines Normalapproximationstheorems nach unten abgeschätzt, anstatt die Tschebyscheff-Ungleichung zu benutzen. Dazu sind einige Vorbemerkungen notwendig.

Zunächst müssen folgende Bedingungen erfüllt sein.

V3.2 a) $\mathfrak{u}$ sei Anfangsstück einer Folge $(u^i)_{i \in \mathbb{N}}$, so daß für je zwei Zustände $z \neq z'$ eine Teilfolge $(u_j^{zz'})_{j \in \mathbb{N}}$ von $(u^i)_{i \in \mathbb{N}}$ existiert mit der Eigenschaft: Für alle $i \in \mathbb{N}$ gibt es ein Ausgabewort v , so daß $p^z(v|u_i^{zz'}) > 0$, $p^{z'}(v|u_i^{zz'}) > 0$ und $p^z(v|u_i^{zz'}) \neq p^{z'}(v|u_i^{zz'})$.

b) $\mathfrak{u}$ "enthält" für alle $z \neq z'$ das Ausgabewort $u_1^{zz'}$.

c) Es gibt eine nur von dem Automaten und von $\mathfrak{u} = (u^1, \dots, u^n)$ abhängige Zahl $\tilde{b}(\mathfrak{u}) > 0$, für die gilt:

Wenn $i \in \{1, \dots, n\}$ und $p^z(v|u^i) > 0$ sowie $p^{z'}(v|u^i) > 0$, dann ist $\left| \ln \frac{p^z(v|u^i)}{p^{z'}(v|u^i)} \right| \leq \tilde{b}(\mathfrak{u})$.⁺

Die Bedingung V3.2 a) sichert, daß $0 < b_0^{zz'}(u_i^{zz'}) < 1$ und zudem $b_j^{zz'}(u_i^{zz'}) > 0$, was unmittelbar einzusehen ist. Zusammen mit V3.2 b) folgt

$$(3.26) \quad \sum_{i=1}^n b_2^{zz'}(u^i) > 0.$$

Aus V3.2 c) und (3.19) findet man

$$(3.27) \quad \sum_{i=1}^n b_3^{zz'}(u^i) \leq \tilde{b}(\mathfrak{u}) \sum_{i=1}^n b_2^{zz'}(u^i).$$

Die ZV $\ln \frac{p^z(\cdot|u^i)}{p^{z'}(\cdot|u^i)}$ ($i=1, \dots, n$) sind statistisch unabhängig, haben nach den auf (3.19) folgenden Bemerkungen existierende dritte Momente und ihre Summe hat einen verschiedenen Erwartungswert. Dann läßt sich das Normalapproximationstheorem von Berry-Esséen (vergl. z.B. [15] p. 288) anwenden.

⁺) Offensichtlich ist V3.2 c) stets erfüllt.

$$(3.28) \left| \mathbb{E}_{p^{22'}(1n)}^{-1} \left\{ \sum_{i=1}^n \ln \frac{\rho^2(\cdot | u^i)}{\rho^{22'}(\cdot | u^i)} < g \left[\sum_{i=1}^n b_2^{22'}(u^i) \right]^{\frac{1}{2}} \right\} - \Phi(g) \right| \\ \leq d \sum_{j=1}^n b_3^{22'}(u^j) / \left[\sum_{i=1}^n b_2^{22'}(u^i) \right]^{\frac{3}{2}},$$

wobei Φ die Verteilungsfunktion der Normalverteilung $\nu_{0,1}$, g eine reelle Zahl und d eine Konstante ist, für die man noch Zolotarev (1967) den Wert 1,322 setzen kann. Aus (3.28) folgt sofort

$$(3.29) \mathbb{E}_{p^{22'}(1n)}^{-1} \left\{ \left| \sum_{i=1}^n \ln \frac{\rho^2(\cdot | u^i)}{\rho^{22'}(\cdot | u^i)} \right| < g \left[\sum_{i=1}^n b_2^{22'}(u^i) \right]^{\frac{1}{2}} \right\} \\ \geq \Phi(g) - \Phi(-g) - 2d \frac{\sum_{i=1}^n b_3^{22'}(u^i)}{\left[\sum_{i=1}^n b_2^{22'}(u^i) \right]^{\frac{3}{2}}}.$$

Wegen (3.26) und (3.27) gilt

$$\frac{\sum_{i=1}^n b_3^{22'}(u^i)}{\left[\sum_{i=1}^n b_2^{22'}(u^i) \right]^{\frac{1}{2}}} \leq \frac{\widehat{b}(v_1)}{\left[\sum_{i=1}^n b_2^{22'}(u^i) \right]^{\frac{1}{2}}}$$

Aus einer Variante der Hölderungleichung (vergl. z.B. [10] p.523) folgt

$$\left[\frac{1}{n} \sum_{i=1}^n b_2^{22'}(u^i) \right]^{\frac{1}{2}} \leq \left[\frac{1}{n} \sum_{i=1}^n b_3^{22'}(u^i) \right]^{\frac{1}{2}},$$

woraus sich ergibt

$$\sum_{i=1}^n b_3^{22'}(u^i) / \left[\sum_{i=1}^n b_2^{22'}(u^i) \right]^{\frac{1}{2}} \geq n^{-\frac{1}{2}}.$$

Schränkt man g auf das Intervall $]0, \sqrt{2}[$ ein, dann ist

$$\Phi(g) - \Phi(-g) \geq 2[2\Phi(\sqrt{2}) - 1]g.$$

Sei mit

$$u_0 := \frac{1 + 2d}{2[2\Phi(\sqrt{2}) - 1]}$$

$$g := \frac{u_0 \widehat{b}(v_1)}{\left[\sum_{i=1}^n b_2^{22'}(u^i) \right]^{\frac{1}{2}}}.$$

Wenn $g < \sqrt{2}$ ist, folgt aus (3.29)

$$(3.30) \quad \mathbb{E}_{p^{z_2'(\cdot|n)}} \left\{ \left| \sum_{i=1}^n \ln \frac{p^z(\cdot|u^i)}{p^{z_2'(\cdot|u^i)}} \right| < \frac{d_0 \tilde{b}(n)}{\left[\sum_{i=1}^n b_2^{z_2'}(u^i) \right]^{\frac{1}{2}}} \left[\sum_{i=1}^n b_2^{z_2'}(u^i) \right]^{\frac{1}{2}} \right\} \geq n^{-\frac{1}{2}}.$$

Im Falle $g \geq \sqrt{2}$ wird wieder Tschebyscheffs Ungleichung angewendet:

$$\mathbb{E}_{p^{z_2'(\cdot|n)}} \left\{ \left| \sum_{i=1}^n \ln \frac{p^z(\cdot|u^i)}{p^{z_2'(\cdot|u^i)}} \right| < d_0 \tilde{b}(n) \right\} \geq \frac{1}{2}.$$

Also gilt für $n \geq 4$ stets die Ungleichung (3.30). Aus (3.16) ergibt sich dann für $n \geq 4$

$$p^z(D_z|n) + p^{z_2'}(D_z|n) \geq n^{-\frac{1}{2}} \exp[-d_0 \tilde{b}(n)] \prod_{i=1}^n b_0^{z_2'}(u^i)$$

und weiterhin, wenn man wie zur Gewinnung von (3.20) vorgeht,

$$\bar{r}(q, \epsilon_d(\cdot|n, \cdot)) \geq \sum_z \sum_{z' \neq z} n^{-\frac{1}{2}} \exp[-\ln|z|^2 - d_0 \tilde{b}(n)] \prod_{i=1}^n b_0^{z_2'}(u^i).$$

Wird

$$(3.31) \quad b_3(n) := -d_0 \tilde{b}(n) - \ln|z|^2$$

gesetzt, so erhält man den folgenden Satz:

S3.7 Satz: Beschreibe $v = (v^1, \dots, v^n)$ die Eingabeworte. v erfülle die Bedingungen V3.1 und V3.2. q^0 sei die Gleichverteilung der Anfangszustände. Für $n \geq 4$ gibt es dann nur von v und dem Automaten abhängige Zahlen $0 < b_3(n) < \infty$ und $0 < s_{zz'}(n) < 1$, so daß

$$(3.32) \quad \bar{p}_{fd}(q, v) \geq \exp[b_3(n) - \ln \sqrt{n}] \sum_z \sum_{z' \neq z} \prod_{i=1}^n \sum_v p^z(v|u^i)^{1-s_{zz'}(n)} p^{z'}(v|u^i)^{s_{zz'}(n)}.$$

Vergleicht man die rechten Seiten von (3.21) und (3.32), dann unterscheiden sich diese nur in dem Exponentialfaktor. Wenn man annimmt, daß $b_3(n)$ in Abhängigkeit von n weniger stark als $\sqrt{n}$ wächst, dann ist (3.32) tatsächlich eine Verschärfung von (3.21) hinsichtlich des asympto-

tischen Verhaltens bzgl. n . Die Annahme wird erfüllt in dem bisher schon mehrfach behandelten Spezialfall.

$\mathcal{W}'$ bestehe wie in 3.1.1 wieder aus Wörtern $u^{z, z'}$, die insgesamt alle Zustände trennen. $\mathcal{W}$ entstehe durch n' -fache periodische Wiederholung von $\mathcal{W}'$ und $(u')_{i \in \mathbb{N}}$ entsprechend. V3.1 und V3.2 seien erfüllt. Durch einfaches Nachprüfen erhält man $b_3(\mathcal{W}) = b_3(\mathcal{W}')$ und, wie schon erläutert wurde, $s_{zz'}(\mathcal{W}) = s_{zz'}(\mathcal{W}')$. Dann folgt für hinreichend große n' mit Hilfe von (3.10)

$$(3.33) \quad \bar{p}_{fd}(q^*, \mathcal{W}) \geq \sum_z \sum_{z' \neq z} \exp \left[b_3(\mathcal{W}') - \ell_n \left[|Z|(|Z|-1)/2 \right]^{\frac{1}{2}} - \ell_n \sqrt{n'} - n' c_o^{zz'}(\mathcal{W}') \right].$$

(3.33) ist also eine Verschärfung von (3.22) für große n' .

In dem hier behandelten Spezialfall läßt sich die Zahl $b_3(\mathcal{W}') - \ell_n \left[|Z|(|Z|-1)/2 \right]^{\frac{1}{2}}$ ohne Schwierigkeiten etwas verbessern. Es wird wieder von (3.29) ausgegangen und gesetzt

$$g := d_o \frac{\sum_{i=1}^n b_3^{zz'}(u')}{\left[\sum_{i=1}^n b_2^{zz'}(u') \right]^{\frac{3}{2}}} = \frac{d_o \sum_{\{z_1, z_2\}} b_3^{zz'}(u^{z_1, z_2})}{\sqrt{n'} \left[\sum_{\{z_1, z_2\}} b_2^{zz'}(u^{z_1, z_2}) \right]^{\frac{3}{2}}}$$

Für hinreichend große n' wird also wegen (3.26) $g < \sqrt{2}$. Definiert man die Zahl $\bar{b}_3(\mathcal{W}')$ durch

$$(3.34) \quad \bar{b}_3(\mathcal{W}') := \max_{z, z' \neq z} \left[-d_o \frac{\sum_{\{z_1, z_2\}} b_3^{zz'}(u^{z_1, z_2})}{\sum_{\{z_1, z_2\}} b_2^{zz'}(u^{z_1, z_2})} + \ell_n \frac{\sum_{\{z_1, z_2\}} b_3^{zz'}(u^{z_1, z_2})}{|Z|^2 \left[\sum_{\{z_1, z_2\}} b_2^{zz'}(u^{z_1, z_2}) \right]^{\frac{3}{2}}} \right],$$

dann erhält man schließlich für hinreichend große n'

$$(3.35) \quad \bar{p}_{fd}(q^0, n) \geq \sum_z \sum_{z' \neq z} \exp \left[\bar{b}_3(n') - \ln \sqrt{n'} - n' c_0^{zz'}(n') \right].$$

Vergleicht man (3.31) und (3.34), dann erkennt man unmittelbar, daß $b_3(n') - \ln [(|Z|-1)/2]^{\frac{1}{2}} \leq \bar{b}_3(n')$

ist, so daß (3.35) eine Verbesserung von (3.33) darstellt.

Eine Abschätzung ohne die Summen, die eine asymptotische Verbesserung von (3.25) ist, erhält man wieder nach Anwendung von Jensens Ungleichung

$$(3.36) \quad \bar{p}_{fd}(q^0, n) \geq \exp \left[c_3(n') - \ln \sqrt{n'} - n' c_0(n') \right],$$

wobei $c_0(n')$ in (3.23) definiert wurde und

$$c_3(n') :=$$

$$\frac{1}{|Z|(|Z|-1)} \sum_z \sum_{z' \neq z} \left[-d_0 \frac{\sum_{\{z_1, z_2\}} b_3^{zz'}(u^{z_1, z_2})}{\sum_{\{z_1, z_2\}} b_2^{zz'}(u^{z_1, z_2})} + \ln \frac{(|Z|-1) \sum_{\{z_1, z_2\}} b_3^{zz'}(u^{z_1, z_2})}{|Z| \left[\sum_{\{z_1, z_2\}} b_2^{zz'}(u^{z_1, z_2}) \right]^{\frac{1}{2}}} \right]$$

3.1.3 Untere Schranken für kleine n

Die Abschätzungen des letzten Abschnittes lieferten im wesentlichen Aussagen über das asymptotische Verhalten des Fehlers $\bar{p}_{fd}(q^0, n)$, wobei q^0 die Gleichverteilung der Anfangszustände ist. In diesem Paragraphen werden Abschätzungen nach unten für $\bar{p}_{fd}(q, n)$ für jedes q angegeben. Diese Abschätzungen haben den Vorteil, daß sie für kleine n relativ gut sind, und den Nachteil, daß sie nicht unmittelbar Auskunft über das asymptotische Verhalten des Fehlers geben. Abschätzungen dieser Art für einfache diagnostische Experimente sind schon in [4] angegeben worden.

Es können hier im Gegensatz zum letzten Abschnitt wieder EV ε_d zugelassen werden, die i.a. nicht deterministisch operieren.

Im folgenden werden diagnostische Vorgabeexperimente als ein Kommunikationsvorgang aufgefaßt. Man stelle sich eine stationäre Quelle vor, die entsprechend der Anfangsverteilung q zu den Zeiten $t=1,2,\dots$ unabhängig Symbole zufällig erzeugt. Der ER wählt zur Zeit $t=0$ ein für alle Mal die Wörter $v=(v^1,\dots,v^n)$. Ein Kanal mit dem Eingabealphabet Z und dem Ausgabealphabet, das aus der Menge aller w mit $\ell(w)=\ell(v)$ besteht und dessen Verhalten allein durch die Wk-maße $p^z(\cdot|v)$ ($z \in Z$) beschrieben wird, überträgt zu den Zeiten $t=1,2,\dots$ unabhängig die von der Quelle erzeugten Symbole z . Der ER würfelt mit $\varepsilon_d(\cdot|v_t)$ unabhängig für jedes vom Kanal ausgegebene w ein $z' \in Z$.

Exakt ausgedrückt, liegt folgender Sachverhalt vor:

Es seien bisher $m \in \mathbb{N}$ Symbole übertragen worden. $\bar{\gamma}$ und $\bar{\lambda}$ seien ZV über $\prod_{t=1}^m Z$, ξ sei ZV über der Menge aller v mit fester Länge $\ell(v)=\ell_1$ und $\bar{\eta}$ sei ZV über alle w_m (in der Bezeichnung aus Kap. 2), wobei w_m aus den Zeilenvektoren w_t' mit $\ell(w_t')=\ell(v)=\ell_1$ ($t=1,\dots,m$) besteht. Die gemeinsame Wk-Verteilung dieser ZV genügt den Bedingungen

$$(3.37) \quad Wk \{ \bar{\lambda}=(z_1',\dots,z_m') | \bar{\eta}=w_m, \bar{\gamma}=(z_1,\dots,z_m), \xi=v \} = \prod_{t=1}^m \varepsilon_d(z_t'|v_t, w_t'),$$

$$(3.38) \quad Wk \{ \bar{\eta}=w_m | \bar{\gamma}=(z_1,\dots,z_m), \xi=v \} = \prod_{t=1}^m p^{z_t}(w_t'|v),$$

$$(3.39) \quad Wk \{ \xi=v | \bar{\gamma}=(z_1,\dots,z_m) \} = p^+(v|z_1),$$

wobei das Funktional $p^+(\cdot)$ dem ER bekannt ist (vergl. die Rolle von $\tilde{p}$ in 2.5),

$$(3.40) \quad Wk \{ \bar{\gamma}=(z_1,\dots,z_m) \} = \prod_{t=1}^m q(z_t).$$

$\bar{\zeta}$ beschreibt also die von der Quelle erzeugten $z \in Z$,
 ξ die Wahl des ER für v , $\bar{\eta}$ die Ausgaben des Automaten
 und $\bar{\lambda}$ die Schätzungen des ER für die von der Quelle er-
 zeugten z .

Eine ZV (α, β, γ) über dem endlichen Raum $A \times B \times C$ wird
 vollständig durch die Zahlen $\text{Wk} \{ \alpha = a, \beta = b, \gamma = c \} = p'(a, b, c)$
 festgelegt. Damit seien wie in der Informationstheorie
 (vergl. z.B. [10] Kap. 2) die folgenden Funktionen defi-
 niert

$$\mathcal{H}(\alpha) := -\mathbb{E} \ln \sum_{b,c} p'(\alpha, b, c),$$

$$\mathcal{H}(\alpha|\beta) := -\mathbb{E} \ln \left[\sum_c p'(\alpha, \beta, c) / \sum_{a,c} p'(a, \beta, c) \right],$$

$$\mathcal{H}(\alpha|\beta, \gamma) := -\mathbb{E} \ln \left[p'(\alpha, \beta, \gamma) / \sum_a p'(a, \beta, \gamma) \right],$$

$$\mathcal{I}(\alpha; \beta) := \mathcal{H}(\alpha) - \mathcal{H}(\alpha|\beta),$$

$$\mathcal{I}(\alpha; \beta|\gamma) := \mathcal{H}(\alpha|\gamma) - \mathcal{H}(\alpha|\beta, \gamma).$$

Mit den bekannten Eigenschaften der Funktionen $\mathcal{H}$ und $\mathcal{I}$
 kann man nun für die in (3.37) bis (3.40) festgelegten ZV
 abschätzen:

$$\mathcal{I}(\bar{\zeta}; \bar{\lambda}|\xi) \leq \mathcal{I}(\bar{\zeta}; \bar{\eta}, \xi) + \mathcal{I}(\bar{\zeta}; \bar{\lambda}|\bar{\eta}, \xi) = \mathcal{I}(\bar{\zeta}; \bar{\eta}, \xi),$$

wobei die letzte Gleichheit wegen (3.37) gilt. Aus (3.39)
 folgt

$$\mathcal{H}(\xi|\bar{\zeta}) = \mathcal{H}(\xi|\zeta_1)$$

mit $\bar{\zeta} = (\zeta_1, \dots, \zeta_m)$. Damit ist

$$\mathcal{I}(\bar{\zeta}; \bar{\eta}, \xi) = \mathcal{I}(\bar{\zeta}; \xi) + \mathcal{I}(\bar{\zeta}; \bar{\eta}|\xi) = \mathcal{I}(\zeta_1; \xi) + \mathcal{I}(\bar{\zeta}; \bar{\eta}|\xi).$$

Wegen (3.38) bis (3.40) gilt

$$\mathcal{H}(\bar{\eta}|\bar{\zeta}, \xi) = \sum_{t=1}^m \mathcal{H}(\eta_t|\zeta_t, \xi),$$

wobei $\bar{\eta}$ die Zeilenvektoren η_t besitzt und

$$\mathcal{K}(\bar{\eta} | \xi) = \sum_{t=1}^m \mathcal{K}(\eta_t | \xi).$$

Daraus schließt man

$$(3.41) \quad \mathcal{J}(\bar{\zeta}; \bar{\eta}, \xi) \leq \mathcal{J}(\zeta_1; \xi) + \sum_{t=1}^m \mathcal{J}(\zeta_t; \eta_t | \xi).$$

Sei

$$\hat{p}_{dm}(v) := \frac{1}{m} \sum_{t=1}^m W_k \{ \eta_t \neq \zeta_t | \xi = v \}.$$

Mit (3.37) bis (3.40) findet man sofort

$$(3.42) \quad \hat{p}_{dm}(v) = \frac{1}{m} \hat{p}_{d1}(v) + \frac{m-1}{m} \bar{r}(q, \varepsilon_d(\cdot | v, \cdot)).$$

Für $k = 2, 3, \dots$ sei, falls $0 < s < 1$,

$$\mathcal{F}_k(s) := s \ln(k-1) - s \ln s - (1-s) \ln(1-s),$$

$$\mathcal{F}_k(0) := 0$$

und

$$\mathcal{F}_k(1) := \ln(k-1).$$

Dann sagt ein bekanntes Lemma (vergl. z.B. [10] Theorem 4.3.2) im hier behandelten Fall aus:

L3.8 Lemma: Falls ξ konstant gleich v ist, gilt

$$(3.43) \quad \mathcal{F}_{|Z|}(\hat{p}_{dm}(v)) \geq \frac{1}{m} \mathcal{K}(\bar{\zeta} | \bar{\eta}, v).$$

Aus (3.40) bis (3.43) erhält man unter Beachtung von

$$\mathcal{J}(\zeta_1; v) = 0$$

$$\mathcal{F}_{|Z|}(\frac{1}{m} \hat{p}_{d1}(v) + \frac{m-1}{m} \bar{r}(q, \varepsilon_d(\cdot | v, \cdot))) \geq \mathcal{K}(\zeta_2) - \frac{1}{m} \sum_{t=1}^m \mathcal{J}(\zeta_t; \eta_t | v).$$

Man schätzt die rechte Seite weiter nach unten ab zu

$$\mathcal{K}(\zeta_2) - \frac{1}{m} \max_{\zeta'_1} \mathcal{J}(\zeta'_1; \eta_1 | v) - \frac{m-1}{m} \max_{\zeta'_2} \mathcal{J}(\zeta_2; \eta_2 | v).$$

$\max_{\zeta'_i}$ bedeutet Maximierung über alle möglichen Verteilungs-

gesetze der Zufallsvariable $\mathcal{Z}'$ über der festen Menge Z . Das Maximum wird bekanntermaßen angenommen. Es ist $\mathcal{J}(\mathcal{Z}', \eta, |v|) \leq \ln |Z|$ und $F_{|Z|}(s)$ eine stetige Funktion von s . Dann erhält man durch Grenzübergang $m \rightarrow \infty$ den Satz S3.9

S3.9 Satz: ν beschreibe die Eingabewörter. $\mathcal{Z}$ sei eine ZV über Z und η eine ZV über der Menge aller w mit $\ell(w) = \ell(\nu)$, deren gemeinsame Verteilung durch die Zahlen $q(z) p^z(w|\nu)$ festgelegt ist. Dann genügt der mittlere erwartete Entscheidungsfehler bei einem diagnostischen Vorgabeexperiment der Ungleichung

$$(3.44) \quad F_{|Z|}(\bar{F}(q, \varepsilon_d(\cdot | \nu, \cdot))) \geq \chi(\mathcal{Z}) - c(\nu)$$

mit

$$c(\nu) := \max_{\mathcal{Z}'} \mathcal{J}(\mathcal{Z}', \eta | \nu).$$

Zunächst werden einige Eigenschaften der Zahlen $c(\nu)$ diskutiert.

1) Wenn $\eta = (\eta^1, \dots, \eta^n)$ und η' ZV über $\{v^i | \ell(v^i) = \ell(u^i)\}$

ist, dann folgt aus der Unabhängigkeit der Reaktionen der Kopien des Automaten

$$(3.45) \quad \mathcal{J}(\mathcal{Z}; \eta^1, \dots, \eta^n | u^1, \dots, u^n) \geq \mathcal{J}(\mathcal{Z}; \eta^1, \dots, \eta^{n-1} | u^1, \dots, u^{n-1}).$$

Ist nun ν Anfang einer Folge $(u^i)_{i \in \mathbb{N}}$, dann existiert

$$\lim_{n \rightarrow \infty} c(\nu).$$

Da am Anfang von Kapitel 3 gefordert wurde, daß der Automat keine äquivalenten Zustände besitzt, gibt es immer Folgen $(u^i)_{i \in \mathbb{N}}$ von Eingabewörtern, die den Voraussetzungen von K3.2 genügen. Wenn ν Anfangsstück einer solchen Folge ist, gilt dann

$$\sup_n c(\nu) = \lim_{n \rightarrow \infty} c(\nu) = \ln |Z|.$$

- 2) Stochastische Automaten haben keinen Vorgriff. Aus dieser Eigenschaft folgt auf ähnliche Weise wie oben:

Wenn $v = (u^1, \dots, u^i, \dots, u^n)$ zu $v^* = (u^1, \dots, u^i, \dots, u^i)$ mit $x \in X$ verlängert wird, ist

$$(3.46) \quad c(v) \leq c(v^*).$$

- 3) Wegen der Unabhängigkeit der Reaktionen der Kopien ist

$$X(\eta^1, \dots, \eta^n | \zeta^1, u^1, \dots, u^n) = \sum_{i=1}^n X(\eta^i | \zeta^1, u^i),$$

$$X(\eta^1, \dots, \eta^n | u^1, \dots, u^n) \leq \sum_{i=1}^n X(\eta^i | u^i).$$

Daraus folgt mit $v = (u^1, \dots, u^n)$

$$c(v) \leq \sum_{i=1}^n c(u^i).$$

Aus der Ungleichung (3.44) lassen sich einige Folgerungen ziehen:

- 4) Wenn $\rho^2(v)$ für ein festes v nicht von z abhängt, dann ist $\mathcal{I}(\zeta^1; \eta(v)) = 0$ für alle $ZV \zeta^1$ über Z . Demnach ist $c(v) = 0$. Liegt nun die Gleichverteilung q^0 der Anfangszustände vor, dann folgt

$$\mathcal{F}_{|Z|}(\bar{v}(q^0, \varepsilon_d(\cdot | v, \cdot))) = \ell_u |Z|.$$

Also ist die Abschätzung (3.44) i.a. nicht verbesserbar.

- 5) Es ist stets $\bar{\rho}_{fd}(q^0, v) \leq \rho_{fd}(v) \leq (|Z|-1)/|Z|$. Wegen der Monotonie von $\mathcal{F}_{|Z|}(s)$ in $0 \leq s \leq (|Z|-1)/|Z|$ folgt dann

$$(3.47) \quad \mathcal{F}_{|Z|}(\rho_{fd}(v)) \geq \ell_u |Z| - c(v).$$

- 6) Sei $0 \leq s \leq (|Z|-1)/|Z|$. Dann läßt sich $\mathcal{F}_{|Z|}(s) - X(\zeta) + c(v)$

in diesem Bereich eindeutig nach s zu einer Funktion $f(|Z|, q, c(v))$ auflösen, die für wachsendes $c(v)$ monoton fällt und für die $f(|Z|, q, X(\zeta)) = 0$ gilt.

Wenn nun $\bar{r}(q, \varepsilon_d(\cdot | v_i)) \leq (|Z|-1)/|Z|$, dann ist

$$(3.48) \quad \bar{r}(q, \varepsilon_d(\cdot | v_i)) \geq \bar{p}_{fd}(q, v) \geq \{(|Z|, q, c(v))\}.$$

7) Sei $|Z| \geq 3$, dann gilt

$$F_{|Z|}(s) \leq s \ln(|Z|-1) + \ln 2$$

und damit folgt aus (3.44) die Vergrößerung

$$\bar{r}(q, \varepsilon_d(\cdot | v_i)) \geq \frac{1}{\ln(|Z|-1)} [\chi(\zeta) - c(v) - \ln 2].$$

8) Sei $v^n := v = (u^1, \dots, u^n)$ Anfang einer Folge $(u^i)_{i \in \mathbb{N}}$, dann ergibt sich aus (3.45)

$$c(v^n) \leq c(v^{n+1}).$$

Sei $0 \leq p_0 \leq \frac{|Z|-1}{|Z|}$ festgelegt. Da die in 6) festgelegte

Funktion $f(|Z|, q, \cdot)$ monoton fällt, gibt es ein größtes $n = n_0(|Z|, q, p_0)$, so daß gilt

$$f(|Z|, q, c(v^n)) \geq p_0.$$

Wählt man dann $n \leq n_0(|Z|, q, p_0)$, so gilt mit (3.48)

$$p_{fd}(v^n) \geq \bar{p}_{fd}(q, v^n) \geq p_0.$$

Damit hat man eine allgemeinere Abschätzung als diejenige in K3.6 erhalten.

9) Wenn $l' \in \mathbb{N}$ ist, dann hängt

$$c'(l') := \max \{c(v) : v = (u^1, \dots, u^n), \max_i l(u^i) = l'\}$$

nur noch von der Länge $l_E := \max_i l(u^i)$ des Experimentes ab. Mit (3.46) ist $(c'(l'))_{l' \in \mathbb{N}}$ monoton

wachsend. Da auch gilt

$$F_{|Z|}(\bar{p}_{fd}(q, v)) \geq \chi(\zeta) - c'(l_E),$$

kann man wie in 8) folgern: Sei p_0 positiv und $p_0 \leq \frac{|Z|-1}{|Z|}$.

Dann gibt es eine Zahl $l_0(|Z|, q, p_0)$ derart, daß

$$p_{fd}(v) \geq \bar{p}_{fd}(q, v) \geq p_0,$$

wenn nur $l_E \leq l_0(|Z|, q, p_0)$.

3.2 Schranken für den Entscheidungsfehler bei einfachen heimführenden Vorgabeexperimenten

In 3.1 wurden n-fach diagnostische Vorgabeexperimente betrachtet. Mit solchen Experimenten wird der Anfangszustand des Automaten, d.h. der Zustand des Automaten vor Beginn des Experimentes, geschätzt. Im vorliegenden Abschnitt sollen Vorgabe-ET untersucht werden, die den Endzustand des Automaten, also denjenigen Zustand des Automaten, der nach der Durchführung des ET erreicht worden ist, bestimmen. Solche ET heißen heimführend (homing). Wie schon in 1.1 angedeutet wurde, sind n-fach heimführende Vorgabe-ET nicht sinnvoll, da die einzelnen Kopien auch bei Eingabe des gleichen Eingabewortes nicht notwendig den gleichen Endzustand erreichen. Es werden daher nur noch einfache heimführende Vorgabe-ET beleuchtet. Damit ist nicht mehr zu erwarten, daß Abschätztechniken, wie sie z.B. in 3.1.1 und 3.1.2 vorgeführt wurden, zu sinnvollen Ergebnissen führen. Allerdings läßt sich die Methode aus 3.1.3 auch auf den hier zu behandelnden Fall übertragen.

Es liege ein Automat mit Anfangsverteilung (X, Y, Z, p, q) vor. Dann sei für alle u und v mit $l(v) = l(u)$

$$(3.49) \quad k_z(v|u) := p_q^z(v|u) := \begin{cases} \sum_{z_0} p(v, z|z_0, u) q(z_0) / \tilde{q}(z|u), & \text{wenn } \tilde{q}(z|u) > 0 \\ \text{beliebig} & \text{sonst} \end{cases}$$

mit

$$(3.50) \quad \tilde{q}(z|u) := \sum_{z_0} \sum_v p(v, z|z_0, u) q(z_0).$$

$p_q^z(v|u)$ ist dann als Wk dafür zu deuten, daß der Automat v ausgibt, wenn u eingegeben wurde, die Anfangsverteilung q vorgelegen hat und der Automat den Zustand z

nach Durchführung des ET erreicht hat. $\tilde{q}(z|u)$ stellt die Wk dafür dar, daß der Automat den Zustand z erreicht hat, wenn u eingegeben wurde und die Anfangsverteilung q vorlag.

A priori Verteilungen $\tilde{q}$ im Sinne von Kap. 2 sind Verteilungen über die Endzustände. Auftreten kann allerdings nur diejenige, die sich wie $\tilde{q}(\cdot|u)$ in (3.50) berechnen läßt.

Eine EV für ein einfaches heimführendes Vorgabe ET werde mit ε_n gekennzeichnet. Bei Benutzung von u als Eingabe ist das bedingte Risiko

$$r(z, \varepsilon_n(\cdot|u, \cdot)) = 1 - \sum_v p_q^z(v|u) \varepsilon_n(z|u, v)$$

und das mittlere Risiko wegen (3.49) und (3.50).

$$\begin{aligned} (3.51) \quad \overline{r}(\tilde{q}(\cdot|u), \varepsilon_n(\cdot|u, \cdot)) &= 1 - \sum_z \sum_v \tilde{q}(z|u) p_q^z(v|u) \varepsilon_n(z|u, v) \\ &= 1 - \sum_z \sum_v \sum_{z_0} q(z_0) p(v, z|z_0, u) \varepsilon_n(z|u, v). \end{aligned}$$

Es gibt nach den Ausführungen in 2.3 Minimax-EV und am wenigsten günstige a priori Verteilungen, doch sind diese im hier behandelten Fall unwichtig, da der ER $\tilde{q}(\cdot|u)$ aus den ihm bekannten p und q berechnen kann. Man interessiert sich daher höchstens noch für Bayes-EV bzgl. $\tilde{q}(\cdot|u)$.

Faßt man ein einfaches heimführendes Vorgabe-EV analog zu den Ausführungen in 3.1.3 als einen Kommunikationsvorgang auf, so liegt folgender Sachverhalt vor: Zur Zeit $t=0$ wählt der ER ein Eingabewort u in Abhängigkeit von p und q des Automaten mit Anfangsverteilung. Eine stationäre Quelle würfele zu den Zeiten $t=1, 2, \dots$ Zustände $z \in Z$

entsprechend dem Wk-maß $\tilde{q}(\cdot|u)$, welches der ER kennt. Ein Kanal mit dem Eingabealphabet Z und dem Ausgabealphabet $\{v: l(v)=l(u)\}$ und dem Übertragungsfunktional $p_q^z(\cdot|u)$ ($z \in Z$) übermittelt die von der Quelle erzeugten Symbole. Der ER würfelt mit einer EV $\varepsilon_h(\cdot|u, \cdot)$ für jedes ankommende v unabhängig ein $z \in Z$, das als Schätzung für den Endzustand des Automaten gelten soll.

Formal bedeutet dies: Die Quelle habe bisher m Symbole erzeugt, der Kanal diese übertragen und der ER entsprechende Entscheidungen getroffen. Seien ξ ZV über X^{l_1} für ein festes $l_1 \in \mathbb{N}$, $\bar{X}$, $\bar{Y}$ und $\bar{N}$ ZV über $\prod_{i=1}^m Z$ und $\bar{v}$ ZV über $\prod_{i=1}^m Y^{l_1}$. Sie genügen den Bedingungen

$$(3.52) \quad Wk\{\bar{N}=(z'_1, \dots, z'_m)\} \cdot \bar{v}=(v_1, \dots, v_m), \bar{X}=(z_1, \dots, z_m), \bar{Y}=(z_1^*, \dots, z_m^*), \xi=u\} \\ = \prod_{t=1}^m \varepsilon_h(z_t | u, v_t),$$

$$(3.53) \quad Wk\{\bar{v}=(v_1, \dots, v_m), \bar{X}=(z_1, \dots, z_m) | \bar{Y}=(z_1^*, \dots, z_m^*), \xi=u\} \\ = \prod_{t=1}^m p(v_t, z_t | z_t^*, u),$$

$$(3.54) \quad Wk\{\bar{Y}=(z_1^*, \dots, z_m^*) | \xi=u\} = \prod_{t=1}^m q(z_t^*) \cdot Wk\{\bar{Y}_1=z_1^* | \xi=u\},$$

wobei $\bar{Y}=(z_1, \dots, z_m)$ gesetzt wurde. $\bar{N}$ beschreibt die Folge der Schätzungen für die Endzustände, $\bar{X}$ die Folge der Endzustände, $\bar{Y}$ diejenige der Anfangszustände, $\bar{v}$ der Ausgaben und ξ die Wahl des ER für das Eingabewort.

Wenn λ_t und χ_t die Komponenten von $\bar{N}$ und $\bar{X}$ sind, dann sei $\hat{p}_{hm}(u) = \frac{1}{m} \sum_{t=1}^m Wk\{\lambda_t \neq \chi_t | \xi=u\}$.

Mit (3.51) bis (3.54) ist dann

$$(3.55) \quad \hat{p}_{hm}(u) = \frac{1}{m} p_{h_1}(u) + \frac{m-1}{m} \bar{r}(\tilde{q}(\cdot|u), \varepsilon_h(\cdot|u, \cdot)).$$

Es gilt mit [10] Theorem 4.3.2 analog zu L3.8

$$(3.56) \quad \mathcal{F}_{|Z|}(\hat{p}_{hm}(u)) \geq \frac{1}{m} \mathcal{X}(\bar{\chi} | \bar{\mathcal{A}}_1, u).$$

Aus (3.52) bis (3.54) folgt

$$(3.57) \quad \mathcal{X}(\bar{\chi} | \bar{\mathcal{A}}_1, u) = \mathcal{X}(\chi_1 | \mathcal{A}_1, u) + (m-1) \mathcal{X}(\chi_2 | \mathcal{A}_2, u).$$

Nach Fortfall des Index 2 erhält man

$$(3.58) \quad \mathcal{X}(\chi | \mathcal{A}, u) = \mathcal{X}(\chi | u) - \mathcal{J}(\chi; \mathcal{A} | u) \\ \geq \mathcal{X}(\chi | u) - \mathcal{J}(\chi; \eta, \zeta | u) - \mathcal{J}(\chi; \mathcal{A} | \eta, \zeta, u) = \mathcal{X}(\chi | \eta, \zeta, u),$$

wobei die letzte Gleichheit aus (3.52) folgt. Mit $m \rightarrow \infty$ ergibt sich dann aus (3.55) bis (3.58) Satz S3.10.

S3.10 Satz: Sei u das Eingabewort für ein einfaches heimführendes Vorgabeexperiment. χ und ζ seien ZV über Z sowie η ZV über $\{v: l(v) = l(u)\}$, deren gemeinsame Verteilung durch die Zahlen $p(v, z | z^*, u) q(z^*)$ vollständig beschrieben wird. Dann genügt der mittlere Entscheidungsfehler der Ungleichung

$$(3.59) \quad \mathcal{F}_{|Z|}(\bar{r}(\tilde{q}(\cdot | u), \varepsilon_n(\cdot | u, \cdot))) \geq \mathcal{X}(\chi | \eta, \zeta, u) \geq \min_{z \in Z} \mathcal{X}(\chi | \eta, z, u)$$

Für drei besondere Automatentypen lassen sich aus (3.59) interessante Ergebnisse folgern.

- 1) Für einen Mealy-Automaten gilt $\text{Wk} \{ \chi = z \mid \eta = v, \zeta = z^*, \xi = u \}$
 $= \text{Wk} \{ \chi = z \mid \zeta = z^*, \xi = u \}$, d.h. Ausgabe und Endzustand sind bedingt unabhängig. Dann gilt:

$$\mathcal{X}(\chi | \eta, \zeta, u) = \mathcal{X}(\chi | \zeta, u).$$

Ist zusätzlich für ein u und alle z^* $W_k\{\chi=z \mid \zeta=z^*, \xi=u\}$ die Gleichverteilung über Z , dann ist $X(\chi \mid \zeta, u) = k_u |Z|$ und in (3.59) gilt das Gleichheitszeichen.

- 2) Ein Automat heißt streng zusammenhängend (indecomposable, vergl. z.B. [10], p. 105), wenn gilt

$$|W_k\{\chi=z \mid \zeta=z_1, \xi=u\} - W_k\{\chi=z \mid \zeta=z_2, \xi=u\}| \xrightarrow{l(u) \rightarrow \infty} 0.$$

Daraus folgt für streng zusammenhängende Mealy-Automaten mit Anfangsverteilung

$$|X(\chi \mid \eta, \zeta, u) - \min_{z \in Z} X(\chi \mid z, u)| \xrightarrow{l(u) \rightarrow \infty} 0.$$

- 3) Ein Automat wird zustandsdeterminiert genannt, wenn $W_k\{\chi=z \mid \zeta=z^*, \xi=u\}$ stets eine 0-1-Verteilung induziert. Da ein zustandsdeterminierter Automat ein Mealy-Automat ist, gilt $X(\chi \mid \eta, \zeta, u) = 0$.

Daraus kann man natürlich nicht schließen, daß für jedes Wort u ein heimführendes Vorgabe-ET mit $\bar{r}=0$ existiert. Man denke dazu nur an deterministische Automaten, für die stets solche u existieren, aber nicht jedes u ein solches ET leistet.

3.3 Bemerkungen

- 1) Zunächst wird ein kleines Beispiel angegeben, das das Verhalten des mittleren Fehlers bei n -fach diagnostischen Vorgabeexperimenten noch einmal beleuchtet. Gegeben sei ein deterministischer Mealy-Automat $(X, Y, Z, \delta, \lambda)$, wobei δ die Zustandsübergänge und λ die Ausgaben beschreibt, d.h. $\delta(z, u)$ gibt den Endzustand an, wenn das Wort u eingegeben wird und z Anfangszustand gewesen ist

und $\lambda(z, u)$ das Ausgabewort unter denselben Bedingungen. Angenommen sei $Z = \{z_1, z_2\}$ und für ein festes Wort u

$\lambda(z_1, u) \neq \lambda(z_2, u)$, d.h. u definiert ein einfaches

diagnostisches Vorgabeexperiment für den deterministischen Automaten (in der Terminologie dieser Arbeit ist dann $p_{fd}(u) = 0$). (X, Y, Z, p) sei ein stochastischer Automat, der "beinahe" wie $(X, Y, Z, \sigma, \lambda)$ arbeitet.

Wenn $0 < t < 1$ und t nahe bei Null ist, sei

$$p^2(v|u) = \begin{cases} 1-t, & \text{wenn } \lambda(z, u) = v \\ \frac{t}{g}, & \text{sonst mit } g := |Y|^{l(u)} - 1 + 1. \end{cases}$$

Gibt man u in n Kopien des stochastischen Automaten, dann kann man bei Vorlage der Gleichverteilung q^0 der Anfangszustände bestenfalls mit einem Entscheidungsfehler $\bar{p}_{fd}(q^0, u)$ mit $v = (u_1, \dots, u_n)$ auf den Anfangszustand schließen. Die Abschätzungen (3.8) und (3.22) geben Auskunft über $\bar{p}_{fd}(q^0, u)$. Dazu wird bemerkt, daß

$$\sum_v p^{z_1}(v|u)^{1-s} p^{z_2}(v|u)^s = (1-t)^{1-s} \frac{t^s}{g^s} + (1-t)^s \frac{t^{1-s}}{g^{1-s}} + \frac{g-1}{g} t.$$

Daraus folgt $s_{z_1 z_2}(u) = \frac{1}{2}$, womit die notwendigen

Konstanten leicht berechnet werden können. Man erhält schließlich

$$\bar{p}_{fd}(q^0, u) \leq \frac{1}{2} \left[2 \frac{(t-t^2)^{\frac{1}{2}}}{g^{\frac{1}{2}}} + \frac{g-1}{g} t \right]^n,$$

$$p_{fd}(q^0, u) \geq \frac{1}{8} \left[2 \frac{(t-t^2)^{\frac{1}{2}}}{g^{\frac{1}{2}}} + \frac{g-1}{g} t \right]^n$$

$$\cdot \exp \left[-\sqrt{n} \frac{2}{\epsilon} \left[-\ln \left[2 \frac{(t-t^2)^{\frac{1}{2}}}{g^{\frac{1}{2}}} + \frac{g-1}{g} t \right] \right]^{\frac{1}{2}} \right].$$

+) Selbstverständlich wird $|Y| \geq 2$ vorausgesetzt.

Diese Abschätzungen liefern zwar das genaue asymptotische Verhalten von $\bar{\rho}_{fa}(q^0, n)$ für $n \rightarrow \infty$, jedoch nicht dasjenige für $t \rightarrow 0$ bei festem n .

- 2) Für deterministische Automaten gilt die Aussage, daß jedes Eingabewort, das ein einfaches diagnostisches Vorgabe-ET leistet, auch ein einfaches heimführendes Vorgabe-ET erlaubt. Dieser Sachverhalt läßt sich i.a. nicht auf stochastische Automaten übertragen. Über die Zusammenhänge dieser beiden Formen von ET bei stochastischen Automaten ist bisher nur wenig bekannt.

- 3) Es sei erwähnt, daß sich die in 3.2 benutzte Schätzmethode für jedes sinnvolle zustandsbestimmende Vorgabe-ET anwenden ließe, um eine untere Schranke für den mittleren Entscheidungsfehler zu gewinnen. Vorausgesetzt wird dabei stets, daß eine Verteilung der Anfangszustände bekannt ist. Die Schätzmethoden aus 3.1.1 bis 3.1.3 können auch bei anderen Arten von n -fach Vorgabe-ET angewendet werden, wenn dabei angenommen wird, daß (in der Terminologie von Kap. 2) die Menge der Kanäle endlich ist, die Aussagenmenge bijektives Bild der Menge der Kanäle ist und die Verlustfunktion die Form $1 - \delta_{ij}$ hat. Es ist nämlich in 3.1.1 bis 3.1.3 die spezielle Struktur von $k_2(\cdot|\cdot) = \rho^2(\cdot|\cdot)$ nicht benutzt worden.

- 4) Zusammenfassend sind in Kap. 3 im wesentlichen folgende Ergebnisse gefunden worden, die von automatentheoretischem Interesse sind: Für n -fach diagnostische Vorgabe-ET ergaben sich bei Vorlage der Gleichverteilung der Anfangszustände bzgl. n asymptotisch scharfe Schranken nach oben und unten für den minimal möglichen mittleren Entscheidungsfehler und bei Vorlage einer

beliebigen Anfangsverteilung Schranken nach oben und unten für denselben Fehler, aus denen aber nicht genau auf das asymptotische Verhalten bzgl. n geschlossen werden kann. Es war möglich, die Minimalzahl der notwendigen Kopien abzuschätzen, damit eine vorgegebene Fehlerschranke unterschritten wird und umgekehrt, die Maximalzahl von Kopien abzuschätzen, so daß eine vorgegebene Fehlerschranke in keinem Falle unterschritten wird. Weiterhin wurde eine obere Schranke für die Länge eines n -fach diagnostischen Vorgabe-ET gefunden, so daß der mittlere Entscheidungsfehler nicht kleiner als eine gegebene Schranke ausfällt. Für einfache heimführende Vorgabe-ET konnte eine untere Schranke für den minimalen mittleren Entscheidungsfehler angegeben werden.

Als wichtigste noch offene Fragen verbleiben für n -fach diagnostische Vorgabe-ET bzgl. n asymptotisch scharfe Schranken des mittleren Entscheidungsfehlers bei beliebigen Anfangsverteilungen und obere Schranken für denselben Fehler, die auch für kleine n gute Abschätzungen liefern. Damit könnte man bei festem n auch die Länge des ET nach unten abschätzen, so daß der Fehler mit Sicherheit kleiner als eine Schranke wird. Für einfache heimführende Vorgabe-ET fehlt noch eine nichttriviale Abschätzung des mittleren Fehlers nach oben.

- 5) Zum Schluß dieser Arbeit ist noch eine Bemerkung angebracht, inwieweit die bisher behandelten Experimente mit stochastischen Automaten aus praktischer Sicht sinnvoll sind. Dazu diene ein Beispiel. Bei einfachen diagnostischen Vorgabe-ET liefert die Ungleichung (3.47) und die Bemerkung 9) aus 3.1.3, daß der mini-

male bedingte Entscheidungsfehler bei Eingabe von beliebigen Worten u in den Automaten stets der Ungleichung

$$\tilde{F}_{|Z|}(p_{fd}(u)) \geq \ln |Z| - \sup_{u'} c'(u') = \ln |Z| - \lim_{u' \rightarrow \infty} c'(u')$$

genügt. Es braucht $\sup c'(u')$ nicht gleich $\ln |Z|$ zu sein. (Dies tritt z.B. schon bei deterministischen Automaten auf, für die kein einfaches diagnostisches Vorgabe-ET existiert, vergl. [4].) In diesem Falle ist es also unmöglich, eine gewisse positive Schranke durch $p_{fd}(u)$ zu unterschreiten. Solche Experimente sind aus praktischer Sicht nicht bedeutsam. Bei n -fach diagnostischen Vorgabe-ET läßt sich der Fehler durch Vergrößerung von n beliebig klein machen. Diese Möglichkeit ist für den praktischen Fall wichtig und stets erwünscht.

Manchmal läßt sich ein Experiment mit einem Automaten beliebig oft wiederholen, z.B. dann, wenn der Automat wieder sicher in seinen Anfangszustand gebracht werden kann. Dann benötigt man nicht n Kopien zum Experimentieren, sondern führt die n Versuche an derselben Kopie aus und kann genau wie beim Arbeiten mit n Kopien schätzen.

Verzeichnis der wichtigsten Definitionen

	Seite
(X, Y, Z, p) stochastischer Automat	1
(X, Y, Z, p, q) stochastischer Automat mit Anfangs- verteilung	2
$(X, Y, Z, \delta, \lambda)$ deterministischer Mealy-Automat	2
K Menge der Kanäle k	8
Vorgriff	9
sequentiell	9
A Menge von Aussagen α	9
n -fach Experiment (ET)	10
ε Experimentiervorschrift (EV)	10
v Verlustfunktion, Fehlerfunktion	11
k Kostenfunktion	12
Entscheidungsproblem	12
r bedingtes Risiko	13
$\bar{r}$ mittleres Risiko	14
$\bar{q}$ a priori Verteilung	14
Bayes-EV bzgl. $\bar{q}$	14
Minimax-EV	14
am wenigsten günstige a priori Verteilung	14
besser als	14
vollständig relativ zu	14
reguläre Topologie in K	15
schwache Topologie in $\mathcal{M}^e(E)$	17
reguläre Topologie in der Menge $\mathcal{E}$ der EV ε	18

	Seite
konvexe Menge von EV	19
$\mathcal{E}^m$ Menge der EV, für die die Wk 1 ist, daß nach m Entscheidungen eine Aussage aus A getroffen ist	25
streng determiniert	30
Abschneiden nach der m -ten Entscheidung	31
buchstabenweise adaptive EV	34
nichtadaptive EV	36
Vorgabe-EV	40
zustandsbestimmendes ET	54
diagnostisches Vorgabe-ET	55
P_{fd} Minimum des maximalen bedingten Ent- scheidungsfehlers	56
$\bar{P}_{fd}$ Minimum des mittleren Entscheidungsfehlers	56
trennen	58
einfach	60
q^0 Gleichverteilung der Anfangszustände	60
deterministisch operieren	62
heimführendes Vorgabe-ET	79
Mealy-Automat	82
streng zusammenhängend	83
zustandsdeterminiert	83

Literaturverzeichnis

- [1] Augustin, U.: Error estimates for low rate codes.
Z. Wahrscheinlichkeitstheorie verw. Geb. 14
(1969), 61-88
- [2] Augustin, U.: Gedächtnisfreie Kanäle für diskrete
Zeit.
Z. Wahrscheinlichkeitstheorie verw. Geb. 6
(1966), 10-61
- [3] Bauer, H.: Wahrscheinlichkeitstheorie und Grundzüge
der Maßtheorie.
De Gruyter, Berlin, 1968
- [4] Böhme, J.F.: Einfache diagnostische Vorgabeexperi-
mente mit stochastischen Automaten.
Erscheint demnächst in den Vortragsveröffent-
lichungen der Tagung "Formale Sprachen und Auto-
matentheorie" in Oberwolfach 1969
- [5] Böhme, J.F.: Über Umformungen von Mustern zur Klassi-
fizierung.
EIK 6 (1970), 51-56
- [6] Booth, T.L.: Estimation, prediction and smoothing
in discrete parameter systems.
Techn. Rpt. Univ. of Connecticut, School of
Eng., Storrs, Connecticut, Dept. of Electrical
Eng., Dec. 1967
- [7] Carlyle, J.W.: On the external probability structure
of finite state channels.
Inf. and Control 7 (1964), 385-397

- [8] Carlyle, J.W.: Reduced forms for stochastic sequential machines.
J. Math. Anal. Appl. 7 (1963), 167-175
- [9] Franz, W.: Topologie I.
Sammlung Götschen, Band 1181, Berlin, 1965
- [10] Gallager, R.G.: Information theory and reliable communication.
J. Wiley, New York, 1968
- [11] Gill, A.: Introduction to the theory of finite-state machines.
Mc Graw-Hill, New York, 1962
- [12] Händler, W.: Einfache diagnostische Experimente bei endlichen Automaten.
3. Coll. über Automatentheorie (Hsg. W. Händler, E. Peschl, H. Unger), Basel, Stuttgart 1967, 56-75
- [13] Knopp, K.: Theorie und Anwendung der unendlichen Reihen.
Springer, Berlin 1947
- [14] Loève, M.: Probability theory.
Van Nostrand, Princeton, 1963
- [15] Neumann, J.v. and Morgenstern, O.: Theory of Games and Economic Behavior.
Princeton Univ. Press, Princeton, 1944
- [16] Paz, A.: Minimization theorems and techniques for sequential stochastic machines.
Inf. and Control 11 (1967), 155 - 166
- [17] Shannon, C.E., Gallager, R.G., Berlekamp, E.R.: Lower bounds to error probability for coding on discrete memoryless channels.
Inf. and Control 10 (1967), 65-103 and 522-552

- [18] Wald, A.: Statistical decision functions.
J. Wiley, New York, 1966
- [19] Wolfowitz, J.: Coding theorems of information
theory.
Springer, Berlin, 1964

Herrn Professor Dr. W. Händler möchte
ich an dieser Stelle für die Anregung
zu dieser Arbeit und ihre großzügige
Förderung danken. Ebenso sei Herrn
Dr. U. Augustin für wertvolle Hinweise
und kritische Bemerkungen gedankt.

Erlangen, den 23.3.70

Johann Friedrich Böhme

Lebenslauf

Am 26. Januar 1940 wurde ich als Sohn des Studienrates Gert Böhme und dessen Frau Gertrud, geb. Holzhausen in Senftenberg geboren. Dort und in Bernau bei Berlin besuchte ich die Grundschule und bis 1958 in Bernau die Oberschule. Nach einer einjährigen Maurerlehre studierte ich die ersten beiden Semester in Dresden Mathematik. Nach nochmaligem Schulbesuch in Delmenhorst zur Anerkennung meines Abiturzeugnisses setzte ich mein Mathematikstudium ab WS 1961/62 an der TH in Hannover fort. Im Dezember 1966 bestand ich die Diplomprüfung. Bis zum Februar 1968 arbeitete ich dann in der Forschungsabteilung der Firma Atlas-Elektronik in Bremen. Anschließend erhielt ich eine Assistentenstelle am Institut für Mathematische Maschinen und Datenverarbeitung der Universität Erlangen-Nürnberg.

Offsetdruck-Fotodruck

J. Hogl, 8520 Erlangen Hauptstr. 109, (Nähe Martin-Luther-Platz) Tel. 22990
